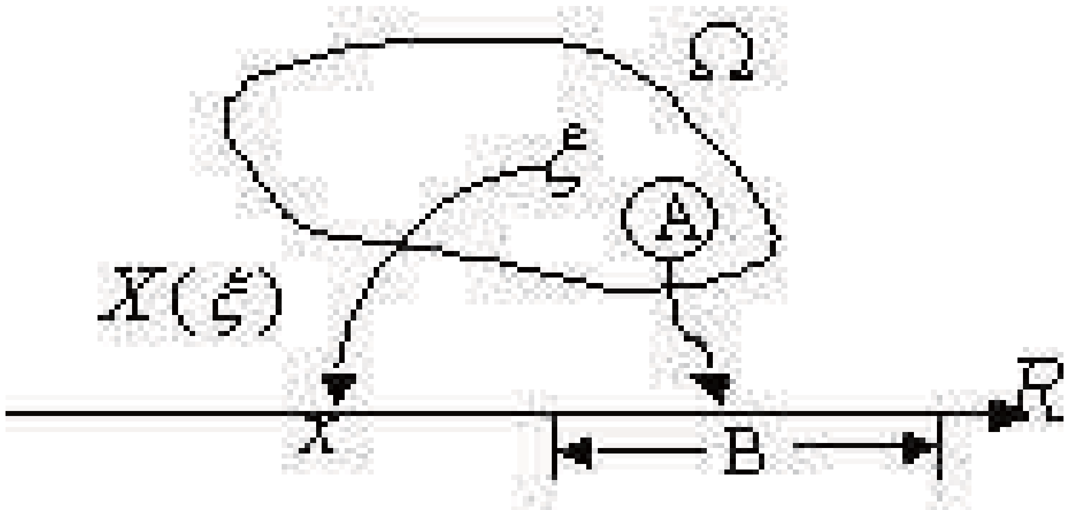
Chap 2.1 : Random Variables

Let Ω be sample space of a probability model, and X a function that maps every $\xi \in \Omega$, to a unique point $x \in R$, the set of real numbers. Since the outcome ξ is not certain, so is the value $X(\xi) = x$. Thus if B is some subset of R, we may want to determine the probability of "$X(\xi) \in B$". To determine this probability, we can look at the set $A = X^{-1}(B) \subset \Omega$. A contains all that maps into B under the function X.



Obviously, if the set $A = X^{-1}(B)$ is an event, the probability of A is well defined; in this case we can say

$$\text{probability of the event "}X(\xi) \in B\text{"} = P(X^{-1}(B)) = P(A)$$

Random Variable (RV): A finite single valued function $X(\cdot)$ that maps the set of all experimental outcomes Ω into the set of real numbers R is said to be a RV.

It is important to identify
the random variable X by the function $X(\xi)$ that maps the sample outcome ξ to the corresponding value of the random variable X. That is

$$\{X = x\} = \{\xi \in \Omega | X(\xi) = x\}$$

Since all events have well defined probability. Thus the probability of the event $\{\xi | X(\xi) \leq x\}$ must depend on x. Denote

$$\boxed{P\{\xi | X(\xi) \leq x\} = F_X(x) \geq 0} \tag{1}$$

The role of the subscript X is only to identify the actual RV. $F_X(x)$ is said to be the Cumulative Distribution Function (CDF) associated with the RV X.

Properties of CDF

- $\boxed{F_X(+\infty) = 1, F_X(-\infty) = 0}$

$$F_X(+\infty) = P\{\xi | X(\xi) \leq +\infty\} = P(\Omega) = 1$$
$$F_X(-\infty) = P\{\xi | X(\xi) \leq -\infty\} = P(\phi) = 0$$

- $\boxed{\text{If } x_1 < x_2, \text{ then } F_X(x_1) \leq F_X(x_2)}$

If $x_1 < x_2$, then the subset $(-\infty, x_1) \subset (-\infty, x_2)$. Consequently the event $\{\xi | X(\xi) \leq x_1\} \subset \{\xi | X(\xi) \leq x_2\}$, since $X(\xi) \leq x_1$, implies $X(\xi) \leq x_2$. As a result

$$F_X(x_1) = P(X(\xi) \leq x_1) \leq P(X(\xi) \leq x_2) = F_X(x_2)$$

implying that the probability distribution function is nonnegative and monotone nondecreasing.

- $\boxed{\text{For all } b > a, F_X(b) - F_X(a) = P(a < X \leq b)}$.

To prove this theorem, express the event $E_{ab} = \{a < X \leq b\}$ as a part of union of disjoint events. Starting with the event $E_b = \{X \leq b\}$. Note that E_b can be written as the union

$$E_b = \{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\} = E_a \cup E_{ab}$$

Note also that E_a and E_{ab} are disjoint so that $P(E_b) = P(E_a) + P(E_{ab})$. Since $P(E_b) = F_X(b)$ and $P(E_a) = F_X(a)$, we can write $F_X(b) = F_X(a) + P(a < X \leq b)$, which completes the proof.

Additional Properties of a CDF

- If $F_X(x_0) = 0$ for some x_0, then $F_X(x)$=0 for $x \leq x_0$.

  This follows, since $F_X(x_0) = P(X(\xi) \leq x_0) = 0$ implies $\{X(\xi) \leq x_0\}$ is the null set, and for any $x \leq x_0$, $\{X(\xi) \leq x\}$ will be a subset of the null set.

- $P\{X(\xi) > x\} = 1 - F_X(x)$

  We have $\{X(\xi) \leq x\} \cup \{X(\xi) > x\} = \Omega$, and since the two events are mutually exclusive, the above equation follows.

- $P\{x_1 < X(\xi) \leq x_2\} = F_X(x_2) - F_X(x_1), x_2 > x_1$

  The events $\{X(\xi) \leq x_1\}$ and $\{x_1 < X(\xi) \leq x_2\}$ are mutually exclusive and their union represents the event $\{X(\xi) \leq x_2\}$.

- $P\{X(\xi) = x\} = F_X(x) - F_X(x^-)$

  Let $x_1 = x - \epsilon, \epsilon > 0$, and $x_2 = x$,

$$\lim_{\epsilon \longrightarrow 0} P\{x - \epsilon < X(\xi) \leq x\} = F_X(x) - \lim_{\epsilon \longrightarrow 0} F_X(x - \epsilon)$$

  or

$$P\{X(\xi) = x\} = F_X(x) - F_X(x^-)$$

$F_X(x_0^+)$, the limit of $F_X(x)$ as $x \to x_0$ from the right always exists and equals $F_X(x_0)$. However the left limit value $F_X(x_0^-)$ need not equal $F_X(x_0)$. $F_X(x)$ need not be continuous from the left. At a discontinuity point of the distribution, the left and right limits are different, and

$$P\{X(\xi) = x_0\} = F_X(x_0) - F_X(x_0^-)$$

Thus the only discontinuities of a distribution function are of the jump type. The CDF is continuous from the right. Keep in mind that the CDF always takes on the upper value at every jump in staircase.

$\boxed{\text{Example 1}}$: X is a RV such that $X(\xi) = c, \xi \in \Omega$. Find $F_X(x)$.

Solution: For $x < c$, $\{X(\xi) \leq x\} = \phi$, so that $F_X(x) = 0$ and for $x \geqslant c$, $\{X(\xi) \leq x\} = \Omega$, so that $F_X(x) = 1$. (see figure below)

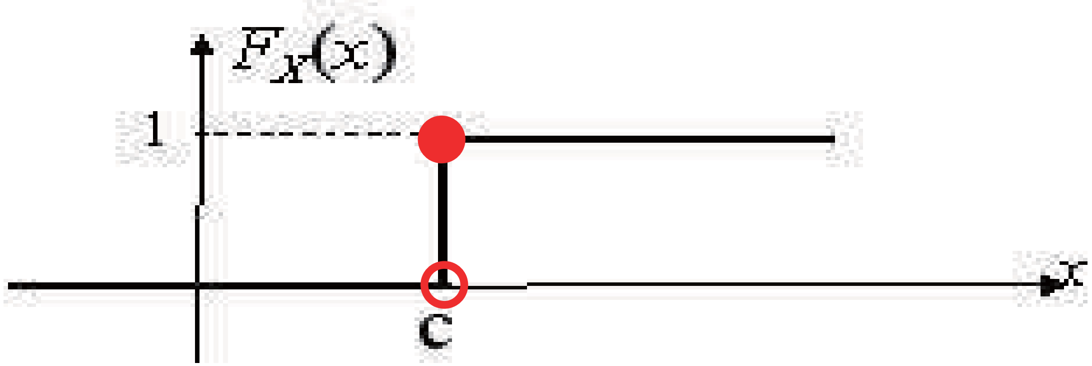


Figure 1: CDF for example 1.

$\boxed{\text{Example 2}}$: Toss a coin. $\Omega = \{H, T\}$. Suppose the RV X is such that X(T)=0, X(H)=1. We know P(T)=q . Find $F_X(x)$.

Solution:

- For $x < 0, \{X(\xi) \leq x\} = \phi$, so that $F_X(x) = 0$.

- For $0 \leq x < 1$, $\{X(\xi) \leq x\} = \{T\}$, so that $F_X(x) = P(T) = q$.
- For $x \geq 1$, $\{X(\xi) \leq x\} = \{H, T\} = \Omega$, so that $F_X(x) = 1$.

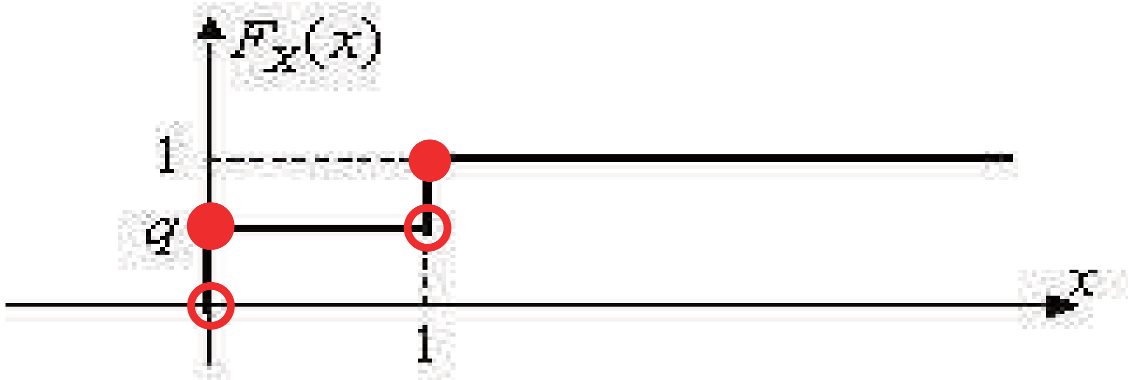


Figure 2: CDF for example 2.

- X is said to be a continuous-type RV if its distribution function $F_X(x)$ is continuous. In that case $F_X(x^-) = F_X(x)$ for all x, therefore, $P\{X = x\} = 0$.
- If $F_X(x)$ is constant except for a finite number of jump discontinuities(piece-wise constant; step-type), then X is said to be a discrete-type RV. If x_i is such a discontinuity point, then

$$p_i = P\{X = x_i\} = F_X(x_i) - F_X(x_i^-)$$

For above two examples, at a point of discontinuity we get

$$P\{X = c\} = F_X(c) - F_X(c^-) = 1 - 0 = 1$$

and

$$P\{X = 0\} = F_X(0) - F_X(0^-) = q - 0 = q$$

Example 3: A fair coin is tossed twice, and let the RV X represent the number of heads. Find $F_X(x)$.

Solution: In this case $\Omega = \{HH, HT, TH, TT\}$, and

$$X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0$$

- $x < 0, \{X(\xi) \leq x\} = \phi \rightarrow F_X(x) = 0$
- $0 \leq x < 1, \{X(\xi) \leq x\} = \{TT\} \rightarrow F_X(x) = P\{TT\} = 1/4$.
- $1 \leq x < 2, \{X(\xi) \leq x\} = \{TT, HT, TH\} \rightarrow F_X(x) = P\{TT, HT, TH\} = \frac{3}{4}$.
- $x \geq 2, \{X(\xi) \leq x\} = \Omega \rightarrow F_X(x) = 1$

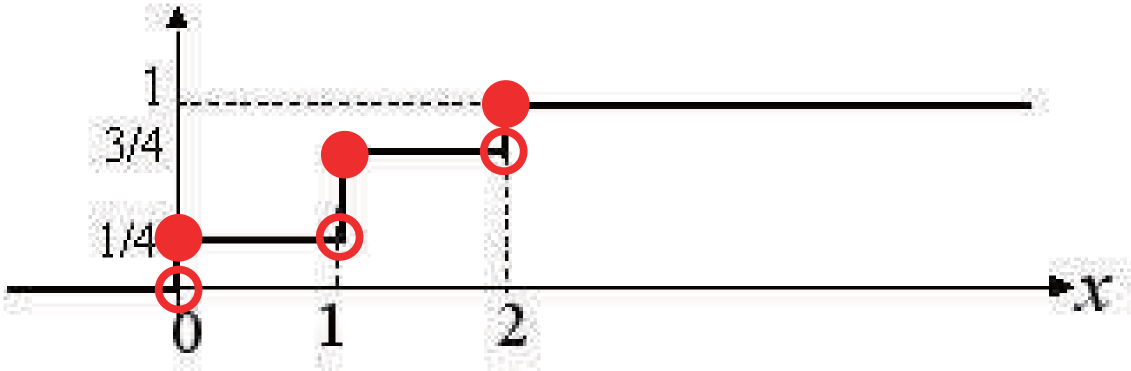


Figure 3: CDF for example 3.

We can also have

P(X=0)=1/4

$$P\{X = 1\} = F_X(1) - F_X(1^-) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$

P(X=2)=1/4

Probability Density Function (pdf)

The first derivative of the distribution function $F_X(x)$ is called the *probability density function* $f_X(x)$ of the RV X. Thus

$$f_X(x) = \frac{d\,F_X(x)}{d\,x}$$

and

$$f_X(x) = \frac{d\,F_X(x)}{d\,x} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \geq 0$$

it follows that $f_X(x) \geq 0$ for all x.

- Discrete RV:if X is a discrete type RV, then its density function has the general form

$$\boxed{f_X(x) = \sum_i p_i \delta(x - x_i)}$$

where x_i represent the jump-discontinuity points in $F_X(x)$. As Fig. 4 shows, $f_X(x)$ represents a collection of positive discrete masses, and it is known as the **probability mass function (pmf)** in the discrete case.

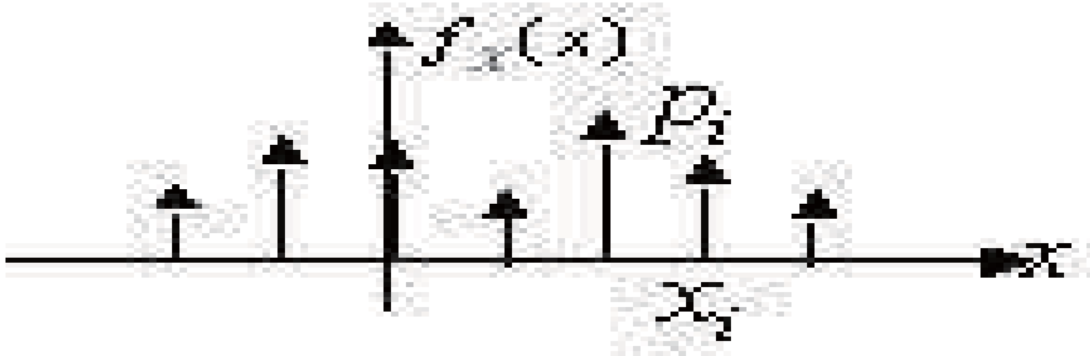


Figure 4: Discrete pmf.

- If X is a continuous type RV, $f_X(x)$ will be a continuous function,
- We also obtain by integration

$$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(u)\,du}$$

Since $F_X(+\infty) = 1$, yields

$$\boxed{\int_{-\infty}^{+\infty} f_X(u)\,du = 1}$$

which justifies its name as the density function.

- we also get (Fig. 5b)

$$\boxed{P\{x_1 < X \le x_2\} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx}$$

Thus the area under $f_X(x)$ in the interval (x_1, x_2) represents the probability in the above equation.

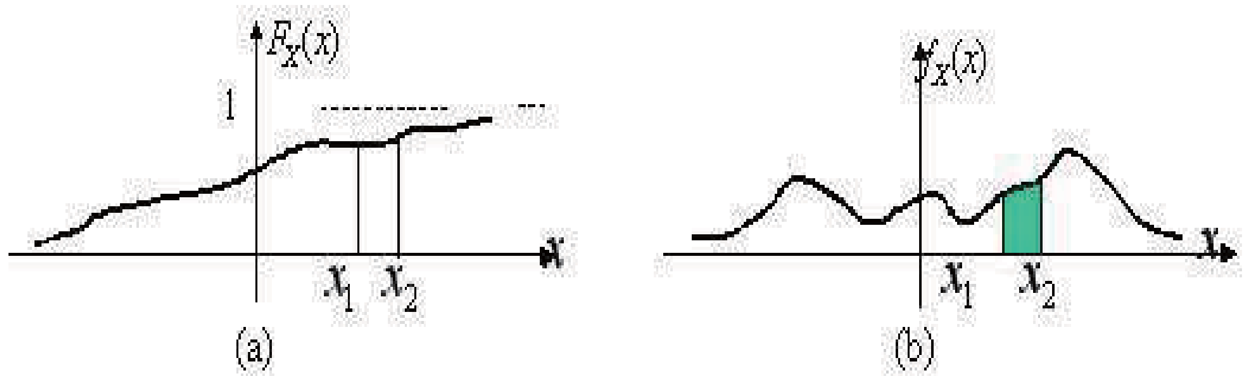


Figure 5: Continuous pdf.

- Often, RVs are referred by their specific density functions - both in the continuous and discrete cases - and in what follows we shall list a number of RVs in each category.

Continuous-type Random Variables

- Normal (Gaussian): X is said to be normal or Gaussian RV, if

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]} \tag{2}$$

This is a bell shaped curve, symmetric around the parameter μ, and its distribution function is given by

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] dy = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)dy$ is called **standard normal CDF**, and is often tabulated.

$$P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = 1 - \Phi(x)$$

$Q(x)$ is called Standard Normal complementary CDF, and $Q(x) = 1 - \Phi(x)$. Since

Table of the Standard Normal Cumulative Distribution Function $\Phi(z)$

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.97725	2.50	0.99379
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.01	0.97778	2.51	0.99396
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.02	0.97831	2.52	0.99413
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.03	0.97882	2.53	0.99430
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.04	0.97932	2.54	0.99446
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.05	0.97982	2.55	0.99461
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.06	0.98030	2.56	0.99477
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.07	0.98077	2.57	0.99492
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.08	0.98124	2.58	0.99506
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.09	0.98169	2.59	0.99520
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	2.10	0.98214	2.60	0.99534
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	2.11	0.98257	2.61	0.99547
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	2.12	0.98300	2.62	0.99560
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	2.13	0.98341	2.63	0.99573
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	2.14	0.98382	2.64	0.99585
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	2.15	0.98422	2.65	0.99598
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	2.16	0.98461	2.66	0.99609
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	2.17	0.98500	2.67	0.99621
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	2.18	0.98537	2.68	0.99632
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	2.19	0.98574	2.69	0.99643
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	2.20	0.98610	2.70	0.99653
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	2.21	0.98645	2.71	0.99664
0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573	2.22	0.98679	2.72	0.99674
0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582	2.23	0.98713	2.73	0.99683
0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591	2.24	0.98745	2.74	0.99693
0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599	2.25	0.98778	2.75	0.99702
0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608	2.26	0.98809	2.76	0.99711
0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616	2.27	0.98840	2.77	0.99720
0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625	2.28	0.98870	2.78	0.99728
0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633	2.29	0.98899	2.79	0.99736
0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641	2.30	0.98928	2.80	0.99744
0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649	2.31	0.98956	2.81	0.99752
0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656	2.32	0.98983	2.82	0.99760
0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664	2.33	0.99010	2.83	0.99767
0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671	2.34	0.99036	2.84	0.99774
0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678	2.35	0.99061	2.85	0.99781
0.36	0.6406	0.86	0.8051	1.36	0.9131	1.86	0.9686	2.36	0.99086	2.86	0.99788
0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693	2.37	0.99111	2.87	0.99795
0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699	2.38	0.99134	2.88	0.99801
0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706	2.39	0.99158	2.89	0.99807
0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713	2.40	0.99180	2.90	0.99813
0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719	2.41	0.99202	2.91	0.99819
0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726	2.42	0.99224	2.92	0.99825
0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732	2.43	0.99245	2.93	0.99831
0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738	2.44	0.99266	2.94	0.99836
0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744	2.45	0.99286	2.95	0.99841
0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750	2.46	0.99305	2.96	0.99846
0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756	2.47	0.99324	2.97	0.99851
0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761	2.48	0.99343	2.98	0.99856
0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767	2.49	0.99361	2.99	0.99861

$\Phi(-z) = 1 - \Phi(z)$

$f_X(x)$ depends on two parameters μ and σ^2, the notation $X \sim N(\mu, \sigma^2)$ is applied. If

$$\boxed{Y = \frac{X-\mu}{\sigma} \quad \sim \quad N(0,1)} \tag{3}$$

Y is called normalized Gaussian RV. Furthermore,

$$aX + b \sim N(a\mu + b, a^2\sigma^2)$$

linear transform of a Gaussian RV is still Gaussian.

- Uniform: $X \sim U(a,b), a < b$, if

$$\boxed{f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{o.w.} \end{cases}} \tag{4}$$

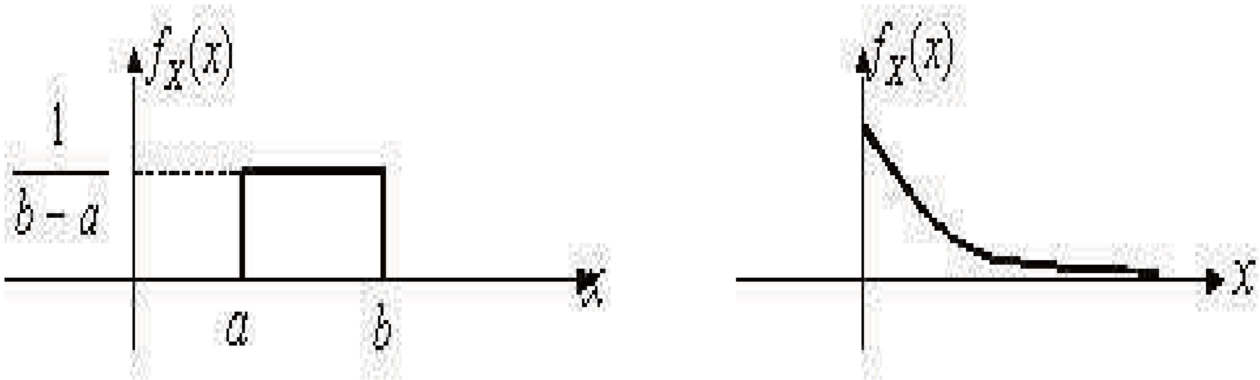


Figure 6: pdf of uniformly distributed and exponential distributed RVs.

- Exponential: $X \sim \epsilon(\lambda)$ if

$$\boxed{f_X(x) = \begin{cases} \frac{1}{\lambda}\exp(-\frac{x}{\lambda}) & x \geq 0 \\ 0 & \text{o.w.} \end{cases}} \tag{5}$$

- Rayleigh $X \sim R(\sigma^2)$

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} & x \geq 0 \\ 0 & \text{o.w.} \end{cases}$$

Let X$=\sqrt{X_1^2 + X_2^2}$ where X_1 and $X_2 \sim N(0, \sigma^2)$ and independent.

Then X is Rayleigh distributed.

Discrete-type Random Variables

- Bernoulli: X takes the values of (0,1), and

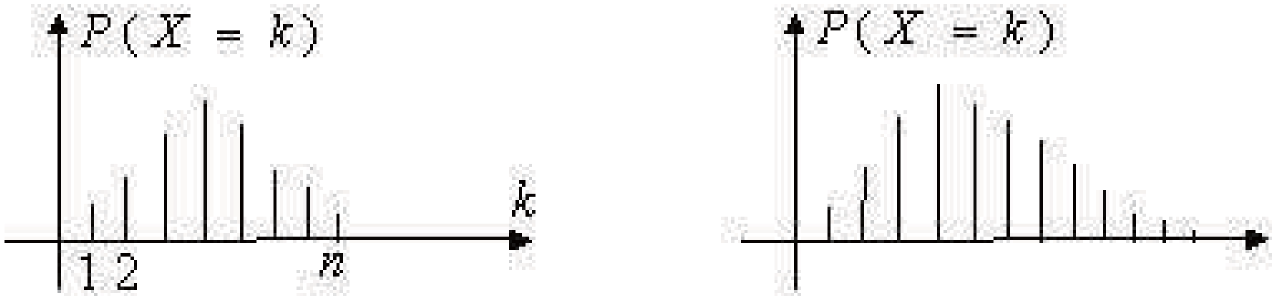
$$P(X=0)=q, \quad P(X=1)=p$$

- Binomial: $X \sim B(n,p)$

$$P(X=k)=\begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}, \quad k=0,1,2,\cdots,n$$

- Poisson: $X \sim P(\lambda)$

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}, \quad k=0,1,2,\cdots,\infty$$



(a) Binomial

(b) Poisson

Figure 7: pmf of Binomial and Poisson distributions.

- Uniform: X takes the values from $[1, \cdots, n]$, and

$$P(X = k) = \frac{1}{n}, \quad k = 1, \cdots, n$$

- Geometric: (number of coin toss till first head appear)

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1, \cdots,$$

where the parameter $p \in (0, 1)$ (probability for head appear on each one toss).

Example of Poisson RV

Example of Poisson Distribution: the probability model of Poisson RV describes phenomena that occur randomly in time. While the time of each occurrence is completely random, there is a known average number of occurrences per unit time. For example, the arrival of information requests at a WWW server, the initiation of telephone call, etc.

For example, calls arrive at random times at a telephone switching office with an average of $\lambda = 0.25$ calls/second. The pmf of the number of calls that arrive in a $T = 2$ second interval is

$$P_K(k) = \begin{cases} (0.5)^k \cdot \frac{e^{-0.5}}{k!} & k = 0, 1, 2, \cdots \\ 0 & \text{o.w.} \end{cases}$$

Example of Binomial RV

Example of using Binomial Distribution: To communicate one bit of information reliably, we transmit the same binary symbol 5 times. Thus, "zero" is transmitted as 00000 and "one" is transmitted as 11111. The receiver detects the correct information if three or more binary symbols are received correctly. What is the information error probability $P(E)$, if the binary symbol error probability is $q = 0.1$?

In this case, we have five trials corresponding to five transmissions. On each trial, the probability of a success is $p = 1 - q = 0.9$ (binary symmetric channel). The error event occurs when the number of successes is strictly less than three:

Let X denote the number of successes out of 5 trials

$$P(E) = P(X=0) + P(X=1) + P(X=2) = q^5 + 5pq^4 + 10p^2q^3 = 0.0081$$

By increasing the number of transmissions (5 times), the probability of error is reduced from 0.1 to 0.0081.

Bernoulli Trial Revisited

Bernoulli trial consists of repeated independent and identical experiments each of which has only two outcomes A or $\overline{A}$ with $P(A) = p$ and $P(\overline{A}) = q$. The probability of exactly k occurrences of A in n such trials is given by Binomial distribution.

Let

$$\boxed{X_k = \text{"exact } k \text{ occurrences in } n \text{ trials"}} \tag{6}$$

Since the number of occurrences of A in n trials must be an integer $k = 0, 1, 2, \cdots, n$, either X_0 or X_1 or X_2 or $\cdots$ or X_n must occur in such an experiment. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = 1 \tag{7}$$

But X_i, X_j are mutually exclusive. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = \sum_{k=0}^{n} P(X_k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \tag{8}$$

From the relation

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

(8) equals $(p+q)^n = 1$, and it agrees with (7).

For a given n and p what is the most likely value of k? The most probable value of k is that number which maximizes in Binomial distribution. To obtain this value, consider the ratio

$$\frac{P_n(k-1)}{P_n(k)} = \frac{n!p^{k-1}q^{n-k+1}}{(n-k+1)!(k-1)!} \cdot \frac{(n-k)!k!}{n!p^k q^{n-k}} = \frac{k}{n-k+1} \cdot \frac{q}{p}$$

Thus $P_n(k) \geq P_n(k-1)$, if $k(1-p) \leq (n-k+1)p$ or $k \leq (n+1)p$. Thus, $P_n(k)$ as a function of k increases until $k=k_m$ where

$$\boxed{k_m = \lfloor (n+1)p \rfloor}$$

Example 4: In a Bernoulli experiment with n trials, find the probability that the number of occurrences of A is between k_1 and k_2.

Solution: with $X_i, i = 0, 1, 2, \cdots, n$ as defined in (6), clearly they are mutually exclusive events. Thus

$$\begin{aligned} P &= P(\text{“Occurrences of } A \text{ are between } k_1 \text{ and } k_2\text{”}) \qquad (9) \\ &= P(X_{k_1} \cup X_{k_1+1} \cup \cdots \cup X_{k_2}) = \sum_{k=k_1}^{k_2} P(X_k) = \sum_{k=k_1}^{k_2} \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} \end{aligned}$$

Example 5: Suppose 5,000 components are ordered. The probability that a part is defective equals 0.1. What is the probability that the total number of defective parts does not exceed 400?

Solution: Let

$$Y_k = \text{“}k \text{ parts are defective among 5000 components”}$$

using (9), the desired probability is given by

$$P(Y_0 \cup Y_1 \cup \cdots \cup Y_{400}) = \sum_{k=0}^{400} P(Y_k) = \sum_{k=0}^{400} \left(\begin{array}{c} 5000 \\ k \end{array} \right) (0.1)^k (0.9)^{n-k}$$

The above equation has too many terms to compute. Clearly, we need a technique to compute the above term in a more efficient manner.

Binomial Random Variable Approximations

Let X represent a Binomial RV, then

$$P(k_1 \leq X \leq k_2) = \sum_{k=k_1}^{k_2} P(X_k) = \sum_{k=k_1}^{k_2} \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} \tag{10}$$

Since the binomial coefficient $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{(n-k)!k!}$ grows quite rapidly with n, it is difficult to compute (10) for large n. In this context, Normal approximation is extremely useful.

Normal Approximation: (Demoivre-Laplace Theorem) Suppose $n \to \infty$ with p held fixed. Then for k in the $\sqrt{npq}$ neighborhood of np, we can approximate

$$\begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right) \tag{11}$$

Thus if k_1 and k_2 in (10) are within or around the neighborhood of the interval

$(np - \sqrt{npq}, np + \sqrt{npq})$ we can approximate the summation in (10) by an integration as

$$P(k_1 \le X \le k_2) = \int_{k_1}^{k_2} \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(x-np)^2}{2npq}\right) dx \tag{12}$$
$$= \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

where

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}$$

We can express (12) in terms of the normalized integral that has been tabulated extensively.

$$erf(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2}\, dy = -erf(-x) \tag{13}$$

$$P(x_1 \le X \le x_2) = erf(x_2) - erf(x_1)$$

Example 6 : A fair coin is tossed 5,000 times. Find the probability that the number of heads is between 2,475 to 2,525.

Solution: We need $P(2475 \leq X \leq 2525)$. Here n is large so that we can use the normal approximation. In this case $p = 1/2$, so that $np = 2500$, and $\sqrt{npq} \approx 35$. Since $np - \sqrt{npq} \approx 2465$ and $np + \sqrt{npq} \approx 2535$, the approximation is valid for $k_1 = 2475$ and $k_2 = 2525$. Thus

$$P(k_1 \leq X \leq k_2) \quad = \quad \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \tag{14}$$

Here

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = -\frac{5}{7} \quad x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{5}{7}$$

Since $x_1 < 0$, from Fig. 8, the above probability is given by

$$P(2475 \leq X \leq 2525) = erf(x_2) - erf(x_1) = erf(x_2) + erf(|x_1|) = 2erf\left(\frac{5}{7}\right) = 0.516$$

where we have used table ($erf(0.7) = 0.258$).

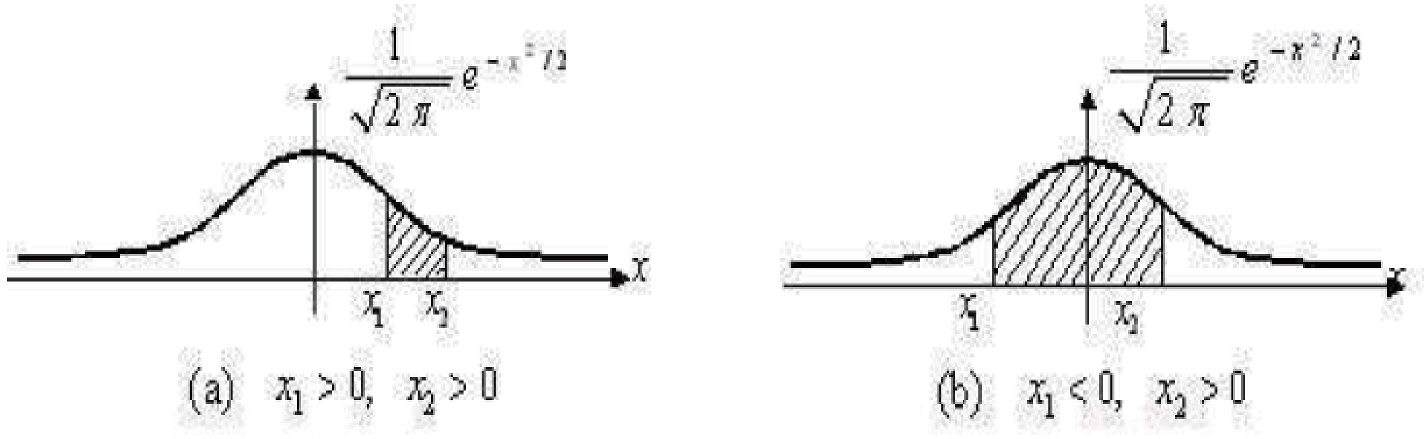


Figure 8: pdf of Gaussian approximation.

Find $P(x_1 \le X \le x_2)$ in Figure 8 using $\Phi(x)$ *or* $erf(x)$ where x is a non-negative number

(a). $P(x_1 \le X \le x_2) = erf(x_2) - erf(x_1) = \Phi(x_2) - \Phi(x_1)$

(b). $P(x_1 \le X \le x_2) = erf(x_2) - erf(x_1) = erf(x_2) + erf(|x_1|)$

$= \Phi(x_2) - \Phi(x_1) = = \Phi(x_2) - (1 - \Phi(|x_1|)) = \Phi(x_2) + \Phi(|x_1|) - 1$

Chap 2.2 : Statistics of RVs

For a RV X, its pdf $f_X(x)$ represents complete information about it. Note that $f_X(x)$ represents very detailed information, and quite often it is desirable to characterize the r.v in terms of its average behavior. In this context, we will introduce two parameters - mean and variance - that are universally used to represent the overall properties of the RV and its pdf.

Mean (Expected Value) of a RV X is defined as

$$\boxed{\overline{X} = E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx} \tag{15}$$

If X is a discrete-type RV, then we get

$$\begin{aligned} \overline{X} &= E(X) = \int x \sum_i p_i \delta(x - x_i)\, dx \\ &= \sum_i x_i p_i = \sum_i x_i P(X = x_i) \end{aligned} \tag{16}$$

Mean represents the average (mean) value of the RV in a very large number of trials. For example

- $X \sim U(a,b)$ (uniform distribution), then,

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\frac{x^2}{2}|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

  is the midpoint of the interval (a,b).

- X is exponential with parameter λ, then

$$E(X) = \int_0^\infty \frac{x}{\lambda}e^{-x/\lambda}\,dx = \lambda\int_0^\infty ye^{-y}\,dy = \lambda \tag{17}$$

  implying that the parameter represents the mean value of the exponential RV.

- X is Poisson with parameter λ, we get

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!} \qquad (18) \\ &= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = \lambda e^{-\lambda}e^{\lambda} = \lambda \end{aligned}$$

  Thus the parameter λ also represents the mean of the Poisson RV.

- X is binomial, then its mean is given by

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} k \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} \qquad (19) \\ &= \sum_{k=1}^{n} k \frac{n!}{(n-k)!k!} p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k} \\ &= np \sum_{i=0}^{n-1} \frac{(n-1)!}{(n-i-1)!\, i!} p^i q^{n-i-1} = np(p+q)^{n-1} = np \end{aligned}$$

Thus np represents the mean of the binomial RV.

- For the normal RV,

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2}\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (y+\mu) e^{-y^2/2\sigma^2}\, dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2}\, dy + \mu \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2}\, dy = \mu \qquad (20) \end{aligned}$$

where the first integral in (20) is zero and the second is 1. Thus the first parameter in $X \sim N(\mu, \sigma^2)$ is in fact the mean of the Gaussian RV X.

Mean of a Function of a RV

Given $X \sim f_X(x)$, suppose $Y = g(X)$ defines a new RV with pdf $f_Y(y)$. Then from the previous discussion, the new RV Y has a mean μ_Y given by

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty} y \, f_Y(y) \, dy \tag{21}$$

From above, it appears that to determine $E(Y)$, we need to determine $f_Y(y)$. However this is not the case if only $E(Y)$ is the quantity of interest. Instead, we can obtain $E(Y)$ as

$$E(Y) = E(g(X)) = \int_{-\infty}^{\infty} y \, f_Y(y) \, dy = \int_{-\infty}^{\infty} g(x) \, f_X(x) \, dx \tag{22}$$

Discrete case

$$E(Y) = \sum_i g(x_i) P(X = x_i) \tag{23}$$

Therefore, $f_Y(y)$ is not required to evaluate $E(Y)$ for $Y = g(X)$. As an example, we determine the mean of $Y = X^2$, where X is a Poisson RV.

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \qquad (24)\\
&= e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{(k-1)!} = e^{-\lambda}\sum_{i=0}^{\infty}(i+1)\frac{\lambda^{i+1}}{i!}\\
&= \lambda e^{-\lambda}\left(\sum_{i=0}^{\infty} i\frac{\lambda^i}{i!} + \sum_{i=0}^{\infty}\frac{\lambda^i}{i!}\right) = \lambda e^{-\lambda}\left(\sum_{i=0}^{\infty} i\frac{\lambda^i}{i!} + e^{\lambda}\right)\\
&= \lambda e^{-\lambda}\left(\sum_{i=1}^{\infty}\frac{\lambda^i}{(i-1)!} + e^{\lambda}\right) = \lambda e^{-\lambda}\left(\sum_{m=0}^{\infty}\frac{\lambda^{m+1}}{m!} + e^{\lambda}\right)\\
&= \lambda e^{-\lambda}(\lambda e^{\lambda} + e^{\lambda}) = \lambda^2 + \lambda
\end{aligned}
$$

In general, $E(X^k)$ is known as the kth moment of RV X. Thus if $X \sim P(\lambda)$, its second moment is $\lambda^2 + \lambda$.

Variance of a RV

Mean alone cannot be able to truly represent the pdf of any RV. As an example to illustrate this, considering two Gaussian RVs $X_1 \sim N(0, 1)$ and $X_2 \sim N(0, 10)$. Both of them have the same mean. However, as Fig. 1 shows, their pdfs are quite different. One is more concentrated around the mean, whereas the other one has a wider spread. Clearly, we need at least an additional parameter to measure this spread around the mean!

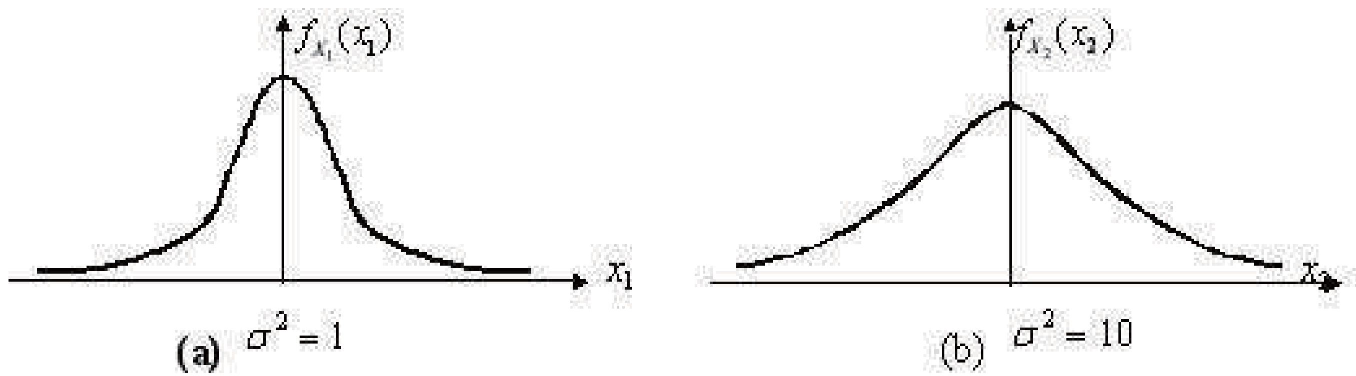


Figure 9: Two Gaussian RV with different variance.

For a RV X with mean μ, $X - \mu$ represents the deviation of the RV from its mean. Since this deviation can be either positive or negative, consider the quantity $(X - \mu)^2$, and its average value $E[(X - \mu)^2]$ represents the average mean square deviation of X around its mean. Define

$$\sigma_X^2 = E[(X - \mu)^2] > 0 \tag{25}$$

With $g(X) = (X - \mu)^2$ and using (22) we get

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)\, dx > 0 \tag{26}$$

σ_X^2 is known as the variance of the RV X, and its square root $\sigma_X = \sqrt{E(X - \mu)^2}$ is known as the standard deviation of X. Note that the standard deviation represents the root mean square spread of the RV X around its mean μ.

Expanding variance definition, and using the linearity of the integrals, we get

$$\begin{aligned} Var(X) &= \sigma_X^2 = \int_{-\infty}^{\infty} (x^2 - 2x\mu + \mu^2) f_X(x)\, dx \qquad (27) \\ &= \int_{-\infty}^{\infty} x^2 f_X(x) dx - 2\mu \int_{-\infty}^{\infty} x f_X(x) dx + \mu^2 \\ &= E(X^2) - \mu^2 = E(X^2) - [E(X)]^2 = \overline{X^2} - \overline{X}^2 \end{aligned}$$

- For a Poisson RV, we can obtain that

$$\sigma_X^2 = \overline{X^2} - \overline{X}^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

  Thus for a Poisson RV, mean and variance are both equal to its parameter λ.

- The variance of the normal RV $N(\mu, \sigma^2)$ can be obtained as

$$Var(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}\, dx \qquad (28)$$

  To simplify the above integral, we can make use of the identity

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}\, dx = 1$$

which gives

$$\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\, dx = \sqrt{2\pi}\sigma$$

Differentiating both sides of above with respect to σ, we get

$$\int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma^3}\, e^{-(x-\mu)^2/2\sigma^2}\, dx = \sqrt{2\pi}$$

or

$$\int_{-\infty}^{\infty} (x-\mu)^2\, \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}\, dx = \sigma^2$$

which represents the $Var(X)$ in (28). Thus for a normal RV $N\ (\mu, \sigma^2)$,

$$Var(X) = \sigma^2$$

therefore the second parameter in $N\ (\mu, \sigma^2)$ in fact represents the variance. As Fig. 9 shows the larger the σ, the larger the spread of the pdf around its mean. Thus as the variance of a RV tends to zero, it will begin to concentrate more and more around the mean, ultimately behaving like a constant.

Moments

As remarked earlier, in general

$$m_n = \overline{X^n} = E(X^n) \quad n \geq 1 \tag{29}$$

are known as the moments of the RV X, and

$$\mu_n = E[(X - \mu)^n]$$

are known as the central moments of X. Clearly, the mean $\mu = m_1$, and the variance $\sigma^2 = \mu_2$. It is easy to relate m_n and μ_n. In fact

$$\begin{aligned} \mu_n &= E((X-\mu)^n) = E\left(\sum_{k=0}^{n} \binom{n}{k} X^k (-\mu)^{n-k}\right) \qquad (30) \\ &= \sum_{k=0}^{n} \binom{n}{k} E(X^k)(-\mu)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} m_k (-\mu)^{n-k} \end{aligned}$$

Direct calculation is often a tedious procedure to compute the mean and variance, and in this context, the notion of the characteristic function can be quite helpful.

Characteristic Function (CF)

The characteristic function of a RV X is defined as

$$\boxed{\Phi_X(\omega) = E(e^{jX\omega}) = \int_{-\infty}^{\infty} e^{jx\omega} f_X(x)\,dx} \tag{31}$$

Thus $\Phi_X(0) = 1$ and $|\Phi_X(\omega)| \leq 1$ for all ω. For discrete RVs the characteristic function is:

$$\boxed{\Phi_X(\omega) = \sum_k e^{jk\omega} P(X = k)} \tag{32}$$

- if $X \sim P(\lambda)$ for poisson distribution, then its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} e^{jk\omega} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{j\omega})^k}{k!} = e^{-\lambda} e^{\lambda\, e^{j\omega}} = e^{\lambda(e^{j\omega}-1)} \tag{33}$$

- if X is a binomial RV, its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{n} e^{jk\omega} \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} (pe^{j\omega})^k q^{n-k} = (pe^{j\omega} + q)^n \tag{34}$$

CF and Moment

To illustrate the usefulness of the characteristic function of a RV in computing its moments, first it is necessary to derive the relationship between them.

$$\begin{aligned} \Phi_X(\omega) &= E(e^{jX\omega}) = E\left[\sum_{k=0}^{\infty} \frac{(j\omega X)^k}{k!}\right] = \sum_{k=0}^{\infty} j^k \, \frac{E(X^k)}{k!} \, \omega^k \qquad (35) \\ &= 1 + jE(X)\,\omega + j^2 \frac{E(X^2)}{2!}\,\omega^2 + \cdots + j^k \frac{E(X^k)}{k!}\,\omega^k + \cdots \end{aligned}$$

where we have used $e^\lambda = \sum_{k=0}^{\infty} \lambda^k/k!$. Taking the first derivative of (35) with respect to ω, and letting it to be equal to zero, we get

$$\frac{\partial \Phi_X(\omega)}{\partial \omega}\,|_{\omega=0} = jE(X) \quad \text{or} \quad E(X) = \frac{1}{j}\,\frac{\partial \Phi_X(\omega)}{\partial \omega}\,|_{\omega=0} \qquad (36)$$

Similarly, the second derivative of (35) gives

$$E(X^2) = \frac{1}{j^2}\,\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2}\,|_{\omega=0} \qquad (37)$$

and repeating this procedure k times, we obtain the kth moment of X to be

$$E(X^k) = \frac{1}{j^k} \frac{\partial^k \Phi_X(\omega)}{\partial \omega^k} |_{\omega=0} \qquad k \geq 1 \tag{38}$$

We can use (35)-(37) to compute the mean, variance and other higher order moments of any RV X.

- if $X \sim P(\lambda)$, then from (33),

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = e^{-\lambda} e^{\lambda e^{j\omega}} \lambda j e^{jw} \tag{39}$$

so that from (36)

$$E(X) = \lambda$$

which agrees with our earlier derivation in (18). Differentiating (39) one more time, we get

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = e^{-\lambda} \left(e^{\lambda e^{j\omega}} (\lambda j\, e^{j\omega})^2 + e^{\lambda e^{j\omega}} \lambda j^2\, e^{j\omega} \right) \tag{40}$$

so that from (37),

$$E(X^2) = \lambda^2 + \lambda$$

which again agrees with results in (24). Notice that compared to the tedious calculations in (18) and (24), the efforts involved by using CF are very minimal.

- We can use the characteristic function of the binomial RV $B(n, p)$ in (34) to obtain its variance. Direct differentiation gives

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = jnp\, e^{j\omega}\,(pe^{j\omega} + q)^{n-1} \tag{41}$$

so that from (36), $E(X) = np$, which is the same as previous calculation.

One more differentiation of (41) yields

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = j^2 np\,[e^{j\omega}\,(pe^{j\omega} + q)^{n-1} + (n-1)p\, e^{j2\omega}\,(p\, e^{j\omega} + q)^{n-2}] \tag{42}$$

and using (37), we obtain the second moment of the binomial r.v to be

$$E(X^2) = np\,(1 + (n-1)\,p) = n^2p^2 + n\,p\,q$$

Therefore, we obtain the variance of the binomial r.v to be

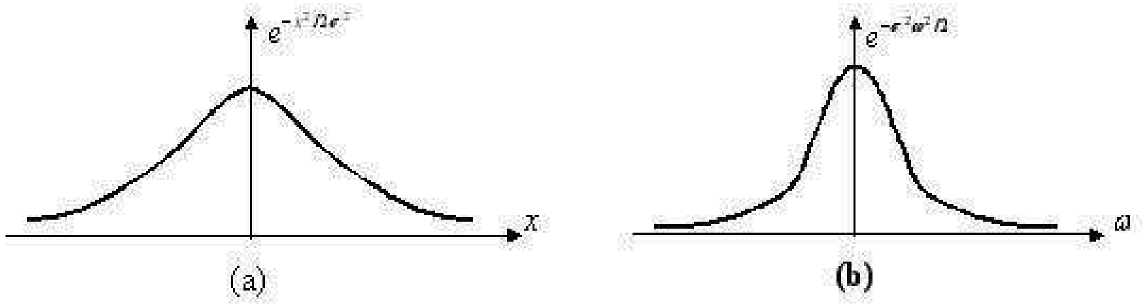
$$\sigma_X^2 = E(x^2) - [E(X)]^2 = n^2p^2 + n\,p\,q - n^2p^2 = n\,p\,q$$

- To obtain the characteristic function of the Gaussian r.v, we can make use of the definition. Thus if $X \sim N(\mu, \sigma^2)$ then

$$\begin{aligned}
\Phi_X(\omega) &= \int_{-\infty}^{\infty} e^{j\omega x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\, dx \quad (\text{let } x - \mu = y) \qquad (43) \\
&= e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{j\omega y}\, e^{-y^2/2\sigma^2}\, dy = e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(y/2\sigma^2)(y - j2\sigma^2\omega)}\, dy \\
&\quad (\text{Let } y - j\sigma^2\omega = \mathrm{z} \text{ so that } y = \mathrm{z} + j\sigma^2\omega) \\
&= e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(\mathrm{z} + j\sigma^2\omega)(\mathrm{z} - j\sigma^2\omega)/2\sigma^2}\, d\mathrm{z} \\
&= e^{j\mu\omega}\, e^{-\sigma^2\omega^2/2} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\mathrm{z}^2/2\sigma^2}\, d\mathrm{z} = e^{(j\mu\omega - \sigma^2\omega^2/2)}
\end{aligned}$$

Notice that the characteristic function of a Gaussian r.v itself has the "Gaussian" bell shape. Thus if $X \sim N(0, \sigma^2)$, then

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \qquad \Phi_X(\omega) = e^{-\sigma^2\omega^2/2}$$



pdf

CF

Figure 10: Gaussian pdf and CF.

From Fig. 10, the reverse roles of σ^2 in $f_X(x)$ and $\Phi_X(\omega)$ are noteworthy $(\sigma^2, vs. 1/\sigma^2)$.

Chebychev Inequality

We conclude this section with a bound that estimates the dispersion of the r.v beyond a certain interval centered around its mean. Since σ^2 measures the dispersion of the RV X around its mean μ, we expect this bound to depend on σ^2 as well.

Consider an interval of width 2ϵ symmetrically centered around its mean μ shown as in Fig. 11. What is the probability that X falls outside this interval? We need

$$P(|X - \mu| \geq \epsilon) = ? \tag{44}$$

To compute this probability, we can start with the definition of σ^2

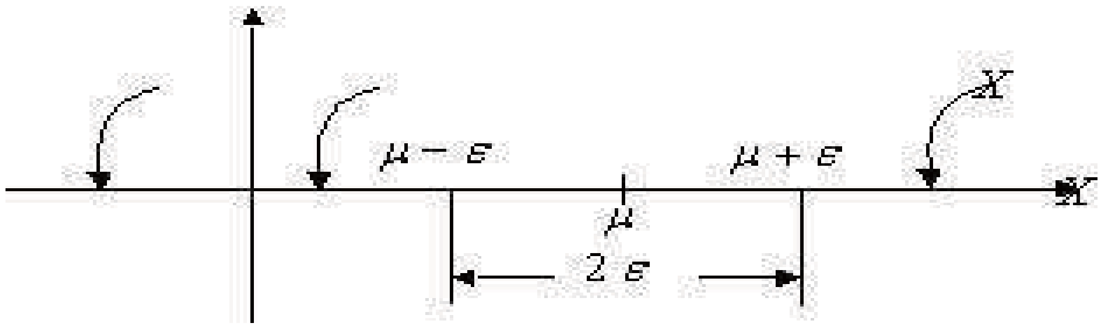


Figure 11: Chebyshev inequality.

$$\begin{aligned}\sigma^2 &= E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx \geq \int_{|x-\mu|\geq\epsilon} (x-\mu)^2 f_X(x)\,dx \quad (45)\\ &\geq \int_{|x-\mu|\geq\epsilon} \epsilon^2 f_X(x)\,dx = \epsilon^2 \int_{|x-\mu|\geq\epsilon} f_X(x)\,dx = \epsilon^2 P(|X-\mu|\geq\epsilon)\end{aligned}$$

From (45), we obtain the desired probability to be

$$\boxed{P(|X-\mu|\geq\epsilon) \leq \tfrac{\sigma^2}{\epsilon^2}} \quad (46)$$

(46) is known as the chebychev inequality. Interestingly, to compute the above probability bound, the knowledge of $f_X(x)$ is not necessary. We only need σ^2, the variance of the RV. In particular with $\epsilon = k\sigma$ in (46) we obtain

$$P(|X-\mu|\geq k\sigma) \leq \frac{1}{k^2} \quad (47)$$

Thus with $k = 3$, we get the probability of X being outside the 3σ interval around its mean to be 0.111 for any RV. Obviously this cannot be a tight bound as it includes all RVs. For example, in the case of a Gaussian RV, from Table $(\mu = 0, \sigma = 1)$:

$$P(|X-\mu|\geq 3\sigma) = 0.0027$$

which is much tighter than that given by (47). Chebychev inequality always underestimates the exact probability.

Example 7: If the height X of a randomly chosen adult has expected value $E[X] = 5.5$ feet and standard deviation $\sigma_X = 1$ foot, use the Chebyshev inequality to find an upper bound on $P(X \geq 11)$

Solution: Since X is nonnegative, the probability that $X \geq 11$ can be written as

$$P[X \geq 11] = P[X - \mu_X \geq 11 - \mu_X] = P[|X - \mu_X| \geq 5.5]$$

Now we use the Chebyshev inequality to obtain

$$P[X \geq 11] = P[|X - \mu_X| \geq 5.5] \leq \frac{Var[X]}{5.5^2} = 0.033 \approx 1/30$$

We can see that the Chebyshev inequality is a loose bound. In fact, $P[X \geq 11]$ is orders of magnitude lower than 1/30. Otherwise, we would expect often to see a person over 11 feet tall in a group of 30 or more people!

Example 8: If X is uniformly distributed over the interval $(0, 10)$, then, as $E[X] = 5$, Var(X)=25/3, it follows from Chebyshev's inequality that

$$P(|X - 5| > 4) \leq \frac{\sigma^2}{\epsilon^2} = \frac{25}{3}\frac{1}{16} \approx 0.52$$

whereas the exact result is

$$P(|X - 5| > 4) = 0.20$$

Thus, although Chebyshev's inequality is correct, the upper bound that it provides is not particularly close to the actual probability.

Similarly, if X is a normal random variable with mean μ and variance σ^2, Chebyshev's inequlity states that

$$P(|X - \mu| > 2\sigma) \leq \frac{1}{4}$$

whereas the actual probability is given by

$$P(|X - \mu| > 2\sigma) = P\left(|\frac{X - \mu}{\sigma}| > 2\right) = 2[1 - \Phi(2)] \approx 0.0456$$

Chebyshev's inequality is often used as a theoretical tool in providing results.

Functions of a Random Variable

Let X be a RV, and suppose $g(x)$ is a function of the variable x. Define

$$Y = g(X)$$

Y is a derived random variable. what is its CDF $F_Y(y)$, pdf $f_Y(y)$?

$\boxed{\text{Example 9}}$: $Y = aX + b$

Solution: Suppose $a > 0$

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

and

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

On the other hand if $a < 0$, then

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X > \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right)$$

and hence

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Therefore, we obtain (for all a)

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Example 10: $Y = X^2$

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) \tag{48}$$

If $y < 0$, then the event $\{X^2 \leq y\} = \phi$, and hence

$$F_Y(y) = 0 \qquad y < 0$$

For $y > 0$, from Fig. 12, the event $\{Y \leq y\} = \{X^2 \leq y\}$ is equivalent to $\{x_1 < X \leq x_2\}$. Hence,

$$\begin{aligned} F_Y(y) &= P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \qquad y > 0 \end{aligned} \tag{49}$$

By direct differentiation, we get

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] & y > 0 \\ 0 & \text{o.w.} \end{cases} \tag{50}$$

If $f_X(x)$ represents an even function, then (50) reduces to

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y})\, U(y) \tag{51}$$

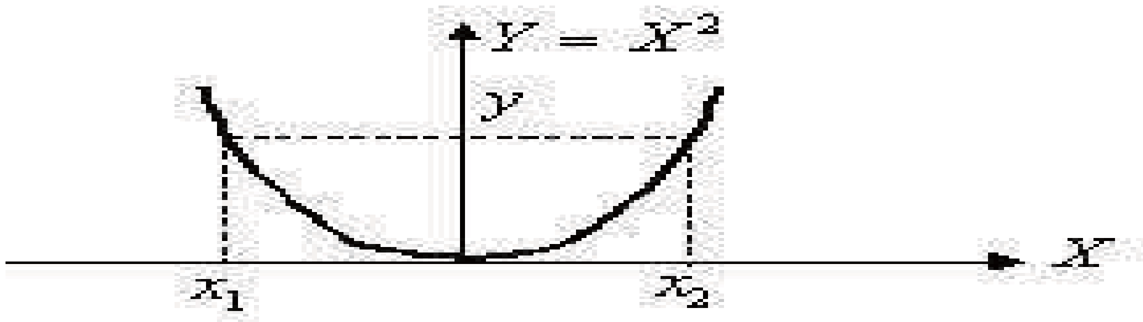


Figure 12: Example $Y = X^2$.

In particular if $X \sim N(0,1)$, so that

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \tag{52}$$

and substituting this into (50) or (51), we obtain the pdf of $Y = X^2$ to be

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}}\, e^{-y/2}\, U(y) \tag{53}$$

General Approach

As a general approach, given $Y = g(X)$, first sketch the graph $y = g(x)$, and determine the range space of y. Suppose $a < y < b$ is the range space of $y = g(x)$.

- for $y < a$, $F_Y(y) = 0$
- for $y > b$, $F_Y(y) = 1$
- $F_Y(y)$ can be nonzero only in $a < y < b$.
- Next, determine whether there are discontinuities in the range space of y. If so evaluate $P(Y(\xi) = y_i)$ at these discontinuities.
- In the continuous region of y, use the basic approach

$$F_Y(y) = P(g(X) \leq y)$$

and determine appropriate events in terms of the RV X for every y. Finally, we must have $F_Y(y)$ for $-\infty < y < +\infty$ and obtain

$$f_Y(y) = \frac{dF_Y(y)}{dy} \quad \text{in} \quad a < y < b$$

However, if $Y = g(X)$ is a continuous function, it is easy to establish a direct procedure to obtain $f_Y(y)$.

Consider a specific y on the y-axis, and a positive increment Δy as shown in Fig. 13.

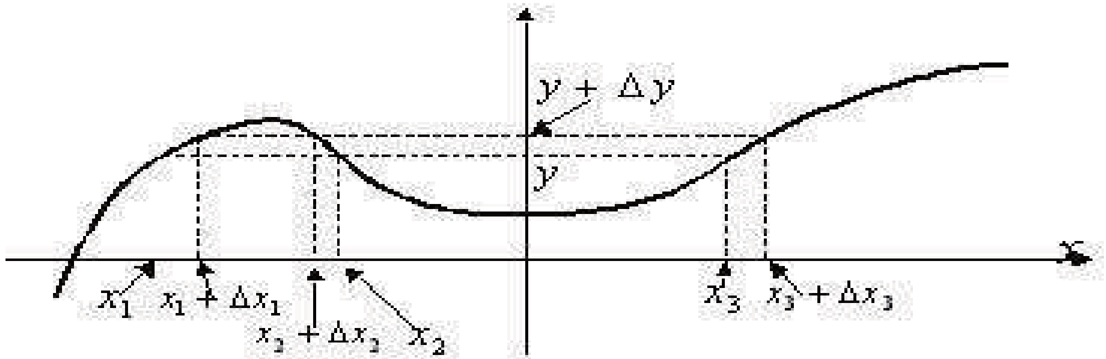


Figure 13: $Y = g(x)$.

$f_Y(y)$ for $Y = g(X)$, where $g(\cdot)$ is of continuous type. we can write

$$P\{y < Y \leq y + \Delta y\} = \int_y^{y+\Delta y} f_Y(u)\,du = f_Y(y) \cdot \Delta y \tag{54}$$

But the event $P\{y < Y \leq y + \Delta y\}$ can be expressed in terms of $X(\xi)$ as well. To see this, referring back to Fig. 13, we notice that the equation $y = g(x)$ has three solutions x_1, x_2, x_3 (for the specific y chosen there). As a result when $\{y < Y \leq y + \Delta y\}$, the RV X could be in any one of the three mutually exclusive intervals

$$\{x_1 < X \leq x_1 + \Delta x_1\} \quad \{x_2+\Delta x_2 < X < x_2\} \quad \{x_3 < X \leq x_3 + \Delta x_3\}$$

Hence the probability of the event in (54) is the sum of the probability of the above three events, i.e.,

$$\begin{aligned} P\{y < Y \leq y + \Delta y\} &= P\{x_1 < X \leq x_1 + \Delta x_1\} + P\{x_2+\Delta x_2 < X < x_2\} \\ &\quad + P\{x_3 < X \leq x_3 + \Delta x_3\} \end{aligned} \tag{55}$$

For small Δy, Δx_i, making use of the approximation in (54), we get

$$f_Y(y)\Delta y = f_X(x_1)\Delta x_1 + f_X(x_2)(-\Delta x_2) + f_X(x_3)\Delta x_3 \tag{56}$$

In this case, $\Delta x_1 > 0$, $\Delta x_2 < 0$ and $\Delta x_3 > 0$ so that (56) can be rewritten as

$$f_Y(y) = \sum_i f_X(x_i)\frac{|\Delta x_i|}{\Delta y} \tag{57}$$

and as $\Delta y \to 0$, (57) can be expressed as

$$f_Y(y) = \sum_i \frac{1}{|dy/dx|_i} f_X(x_i) = \sum_i \frac{1}{|g'(x_i)|_i} f_X(x_i) \tag{58}$$

The summation index i in (58) depends on y, and for every y the equation $y = g(x_i)$ must be solved to obtain the total number of solutions at every y, and the actual solutions $x_1, x_2, \cdots$ all in terms of y.

For example, if $Y = X^2$, then for all $y > 0, x_1 = -\sqrt{y}$ and $x_1 = \sqrt{y}$ represent the two solutions for each y. Notice that the solutions x_i are all in terms of y so that the right side of (58) is only a function of y. Referring back to the example $Y = X^2$ here for each $y > 0$, there are two solutions given by $x_1 = -\sqrt{y}$ and $x_2 = +\sqrt{y}$ ($f_Y(y) = 0$ for $y < 0$). Moreover

$$\frac{dy}{dx} = 2x \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x=x_i} = 2\sqrt{y}$$

and using (58) we get

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] & y > 0 \\ 0 & \text{o.w.} \end{cases} \tag{59}$$

which agrees with earlier result.

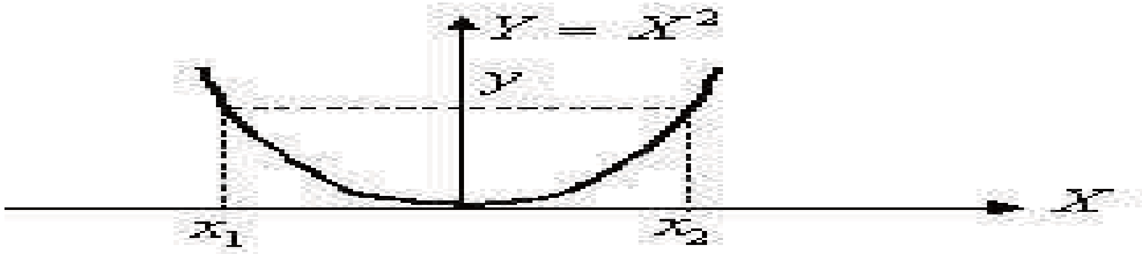


Figure 14: $Y = X^2$.

$\boxed{\text{Example 11}}$: Let $Y = 1/X$, find $f_Y(y)$.

Solution: Here for every $y, x_1 = 1/y$ is the only solution, and

$$\frac{dy}{dx} = -\frac{1}{x^2} \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x=x_1} = \frac{1}{1/y^2} = y^2$$

and substituting this into (58), we obtain

$$f_Y(y) = \frac{1}{y^2} f_X\left(\frac{1}{y}\right). \tag{60}$$

In particular, suppose X is a Cauchy r.v with parameter α so that

$$f_X(x) = \frac{\alpha/\pi}{\alpha^2 + x^2} \qquad -\infty < x < \infty$$

In that case from (60), $Y = 1/X$ has the pdf

$$f_Y(y) = \frac{1}{y^2} \cdot \frac{\alpha/\pi}{\alpha^2 + (1/y)^2} = \frac{(1/\alpha)/\pi}{(1/\alpha)^2 + (y)^2} \qquad -\infty < x < \infty$$

Functions of A Discrete-type RV

Suppose X is a discrete-type RV with

$$P(X = x_i) = p_i, \quad x = x_1, x_2, \cdots, x_i, \cdots,$$

and $Y = g(X)$. Clearly Y is also of discrete-type, and when $x = x_i, y_i = g(x_i)$, and for those y_i,

$$P(Y = y_i) = P(X = x_i) = p_i, \quad y = y_1, y_2, \cdots, y_i, \cdots \tag{61}$$

Example 12 : Suppose $X \sim P(\lambda)$, so that

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad k = 0, 1, 2, \cdots$$

Define $Y = X^2 + 1$. Find the pmf of Y.

Solution: X takes the values $0, 1, 2, \cdots, k, \cdots$, so that Y only takes the values $1, 2, 5, \cdots, k^2 + 1, \cdots$,

$$P(Y = k^2 + 1) = P(X = k)$$

so that for $j = k^2 + 1$

$$P(Y = j) = P(X = \sqrt{j-1}) = e^{-\lambda}\frac{\lambda^{\sqrt{j-1}}}{(\sqrt{j-1})!}, \quad j = 1, 2, 5, \cdots, k^2+1, \cdots$$