The **probability mass func** of X is $f(x) = P(X = x)$, and satisfies

$$f(x) > 0, x \in S \quad \text{and} \quad \sum_{x \in S} f(x) = 1 \quad \text{and} \quad P(X \in A) = \sum_{x \in A} f(x)$$

The **cumulative distribution func** of X is $F(x) = P(X \le x)$
The **mathematical expectation** (expect value) of $u(X)$ is
$E[u(X)] = \sum_{x \in S} u(x) f(x)$
The **mean** of X is $\mu = E(X)$
The **variance** of X is $\sigma^2 = E[(X-\mu)^2] = E(X^2) - \mu^2$
The r-th **moment** of X about b is $E[(X-b)^r]$
The **moment-generating func** of X is $M(t) = E(e^{tX})$
$M'(0) = E(X), M''(0) = E(X^2)$

hypergeometric distribution
choose n chips from N_1 red chips and N_2 blue chips
X = # of red chips

$$f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}$$

$$\mu = n\binom{N_1}{N}, \quad \sigma^2 = np(1-p)\binom{N-n}{N-1}$$

binomial distribution: $b(n,p)$
X = # of success in n bernoulli trials

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad M(t) = (q + pe^t)^n$$

$$\mu = np, \quad \sigma^2 = np(1-p)$$

when $n = 1$, it is equivalent to **Bernoulli distribution**

negative binomial distribution
X = # of success to observe until r-th success

$$f(x) = \binom{x-1}{r-1} p^r q^{x-r}, \quad M(t) = p^r[1 - qe^t]^{-r}$$

$$\mu = \frac{r}{p}, \quad \sigma^2 = \frac{r(1-p)}{p^2}$$

when $r = 1$, it is equivalent to **geometric distribution**

posisson distribution
X = # of occurences of event in unit interval with mean of occurences
$\lambda = E(X) > 0$
the probability of exactly one occurences in interval of length h is $p = h\lambda$

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad M(t) = e^{\lambda(e^t - 1)}$$

$$\mu = \sigma^2 = \lambda$$

The **probability density func** of X is $f(x) = F'(x)$ (if exist), and satisfies

$$f(x) \ge 0, x \in S \quad \text{and} \quad \int_S f(x)\, dx = 1 \quad \text{and} \quad P(a < X < b) = \int_a^b f(x)\, dx$$

The **cumulative distribution func** of X is $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt$

The **mathematical expectation** (expect value) of $u(X)$ is
$E[u(X)] = \int_{-\infty}^{\infty} u(x) f(x)\, dx$

gamma distribution
W = waiting time until α-th occurrence
λ is the mean of occurences in unit interval, and $\theta = 1/\lambda$ is the mean waiting time for the first occurence

$$F(w) = P(W \le w) = 1 - P(W > w) = 1 - \sum_{k=0}^{\alpha-1} \frac{(\lambda w)^k e^{-\lambda w}}{k!}$$

$$f(w) = \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!} e^{-\lambda w} = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, \quad M(t) = \frac{1}{(1-\theta t)^\alpha}$$

$$\mu = \alpha\theta, \quad \sigma^2 = \alpha\theta^2$$

gamma function is $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y}\, dy = (t-1)!$
when $\alpha = 1$, it is equivalent to **exponential distribution**
when $\alpha = r/2$ and $\theta = 2$, we say X has a **chi-square distribution** $\chi^2(r)$ with r degrees of freedom, and we have

$$P[X \ge \chi^2_k(r)] = k, \quad P[X \le \chi^2_{1-k}(r)] = k$$

Example: If customers arrive at a shop on the average of 30 per hour, what is the probability that the shopkeeper will have to wait longer than 9.390 minutes for the first nine customers to arrive? the mean rate of arrivals per minute is $\lambda = 1/2$. Thus, $\theta = 2$ and $\alpha = 9 = r/2$. If X denotes the waiting time until the ninth arrival, then X is $\chi^2(18)$.

$$P(X > 9.390) = 1 - 0.05 = 0.95$$

normal distribution: $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad M(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

if Z is $N(0,1)$, we say Z has a **standard normal distribution**, and the cdf of Z is

$$P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw, \quad P(Z > z) = \Phi(-z) = 1 - \Phi(z)$$

If X is $N(\mu, \sigma^2)$, then $Z = \frac{(X-\mu)}{\sigma}$ is $N(0,1)$.
If X is $N(\mu, \sigma^2)$, then $V = \frac{(X-\mu)^2}{\sigma^2} = Z^2$ is $\chi^2(1)$.

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

$$\sum_{n=0}^{\infty} \binom{n}{k} x^k = (1+x)^n$$

$$\int \ln(x)\, dx = x\ln(x) - x + C$$

$$\int_a^b f(g(x))g'(x)\, dx = \int_{g(a)}^{g(b)} f(u)\, du,$$
where $u = g(x)$ and $du = g'(x)\, dx$

$$\int f(x)g'(x)\, dx = f(x)g(x) - \int f'(x)g(x)\, dx$$

$$\frac{d}{dx}\, a^x = \ln(a)a^x$$

$$\frac{d}{dx}\, e^{g(x)} = e^{g(x)} g'(x)$$

$$\frac{d}{dx}\, \log_a(x) = \frac{1}{\ln(a)x}$$

$$\frac{d}{dx}\, \ln(x) = \frac{1}{x}$$

$$\frac{d}{dx}\, f(x)g(x) = f'(x)g(x) + f(x)g'(x)$$

$$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2}$$

$$\frac{d}{dx}\, f(g(x)) = f'(g(x)) \cdot g'(x)$$

The **joint probability mass func** of X, Y is $f(x,y) = P(X=x, Y=y)$, and statisfies

$$0 \le f(x,y) \le 1 \quad \text{and} \quad \sum\sum_{(x,y)\in S} f(x,y) = 1$$

$$\text{and} \quad P[(X,Y)\in A] = \sum\sum_{(x,y)\in A} f(x,y)$$

The **marginal probability mass func** of X is
$f_X(x) = \sum_y f(x,y) = P(X=x), \quad x \in S_X$

X and Y is independent if and only if $f(x,y) = f_X(x) f_Y(y)$

The **expected value** of $u(X,Y)$ is $E[u(X,Y)] = \sum\sum_{(x,y)\in S} u(x,y) f(x,y)$

The **mean & variance** of X is

$$\mu_X = E(X) = \sum_X x f_X(x), \quad \sigma_X^2 = E[X^2] - \mu_X^2 = \left[\sum_X x^2 f_X(x)\right] - \mu_X^2$$

The **covariance** of X and Y is
$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X \mu_Y$

The **correlation coefficient** of X and Y is $\rho = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$

$$E(XY) = \mu_X \mu_Y + \rho \sigma_X \sigma_Y$$

expected value of squared distance between each (x,y) and line $y = \mu_Y + b(x - \mu_X)$ is
$E\{[(Y-\mu_Y) - b(X-\mu_X)]^2\} = K(b) = \sigma_Y^2 - 2b\rho\sigma_X\sigma_Y + b^2\sigma_X^2$
where $K(b)$ obtain minimum at $b = \rho\sigma_Y/\sigma_X$
and the **least squares regression line** is $y = \mu_Y + \rho\dfrac{\sigma_Y}{\sigma_X}(x - \mu_X)$

Conditional Distribution

The **conditional probability mass function** of Y, given that $X = x$, is
$h(y|x) = \dfrac{f(x,y)}{f_X(x)}$, and $P(a < Y < b \mid X = x) = \sum_{y: a<y<b} h(y|x)$

The **conditional mean** of Y is $\mu_{Y|x} = E(Y|x) = \sum_y y h(y|x)$

The **conditional variance** of Y is
$\sigma_{Y|x}^2 = E\{[Y - E(Y|x)]^2 \mid x\} = E(Y^2|x) - [E(Y|x)]^2$

If $E(Y|x) = a + bx$ (a linear function of x), then
$E(Y|x) = \mu_y + \rho\dfrac{\sigma_Y}{\sigma_X}(x - \mu_X)$

Continuous Type

The **joint probability density func** of X, Y is $f(x,y) = P(X=x, Y=y)$, and statisfies

$$f(x,y) > 0 \quad \text{and} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

$$\text{and} \quad P[(X,Y)\in A] = \int\int_A f(x,y)\,dx\,dy$$

The **marginal probability density func** of X is
$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = P(X=x), \quad x \in S_X$

Bivariate Normal Distribution

Let $f(x,y)$ be the joint pdf of continuous rv X and Y, if

- X has normal distribution, that is $f_X(x) = \dfrac{1}{\sigma_X\sqrt{2\pi}} \exp\left[-\dfrac{(x-\mu_X)^2}{2\sigma_X^2}\right]$
- conditional distribution of Y is normal: $N\left[E(Y|x), \sigma_{Y|x}^2\right]$ and

$h(y|x) = \dfrac{1}{\sigma_{Y|x}\sqrt{2\pi}} \exp\left[-\dfrac{(y-\mu_Y)^2}{2\sigma_{Y|x}^2}\right]$

- conditional mean of Y is linear function of x, that is
$E(Y|x) = \mu_y + \rho\dfrac{\sigma_Y}{\sigma_X}(x - \mu_X)$
- conditional variance of Y: $\sigma_{Y|x}^2 = \sigma_Y^2(1-\rho^2)$ is constant

then $f(x,y)$ is called **bivariate normal pdf**, and

$$f(x,y) = h(y|x) f_X(x) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{q(x,y)}{2}\right]$$

$$q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

Let us assume that in a certain population of college students, the respective grade point averages—say, X and Y—in high school and the first year in college have an approximate bivariate normal distribution with parameters $\mu_X = 2.9$, $\mu_Y = 2.4$, $\sigma_X = 0.4$, $\sigma_Y = 0.5$, and $\rho = 0.8$.

Then, for example,

$$P(2.1 < Y < 3.3) = P\left(\frac{2.1-2.4}{0.5} < \frac{Y-2.4}{0.5} < \frac{3.3-2.4}{0.5}\right)$$
$$= \Phi(1.8) - \Phi(-0.6) = 0.6898.$$

Since the conditional pdf of Y, given that $X = 3.2$, is normal with mean

$$2.4 + (0.8)\left(\frac{0.5}{0.4}\right)(3.2 - 2.9) = 2.7$$

and standard deviation $(0.5)\sqrt{1-0.64} = 0.3$, we have

$$P(2.1 < Y < 3.3 \mid X = 3.2)$$
$$= P\left(\frac{2.1-2.7}{0.3} < \frac{Y-2.7}{0.3} < \frac{3.3-2.7}{0.3} \,\middle|\, X = 3.2\right)$$
$$= \Phi(2) - \Phi(-2) = 0.9544.$$

Theorem 4.5-1

If X and Y have a bivariate normal distribution with ρ, then X and Y are independent if and only if $\rho = 0$.

Functions of Random Variables

Let $Y = u(X)$ be a function of X, how to find distribution of Y:

1. Distribution function technique

If we can find it's cdf $G(y) = P(Y \le y) = P[u(X) \le y]$, then it's pdf $g(y) = G'(y)$.

loggamma pdf

X has a gamma distribution, and $Y = e^X$

$G(y) = P(X \le \ln y) = \int_0^{\ln y} \dfrac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}\,dx$

$g(y) = \dfrac{1}{\Gamma(\alpha)\theta^\alpha}\dfrac{(\ln y)^{\alpha-1}}{y^{1+1/\theta}}, \quad 1 < y < \infty$

2. Change-of-variable technique

X is a continuous rv with pdf $f(x)$ and $c_1 < x < c_2$. Taking $Y = u(X)$ as a continuous *increasing/decreasing* function with inverse function $X = v(Y)$.

$$g(y) = f[v(y)]|v'(y)|, \quad d_1 = u(c_1) < y < d_2 = u(c_2)$$

Theorem 5.1-1

Let Y have a distribution that is $U(0,1)$. Let $F(x)$ have the properties of a cdf of the continuous type with $F(a) = 0, F(b) = 1$, and suppose that $F(x)$ is strictly increasing on the support $a < x < b$, where a and b could be $-\infty$ and ∞, respectively. Then the random variable X defined by $X = F^{-1}(Y)$ is a continuous-type random variable with cdf $F(x)$.

Theorem 5.1-2

Let X have the cdf $F(x)$ of the continuous type that is strictly increasing on the
support $a < x < b$. Then the random variable Y, defined by $Y = F(X)$, has a distribution that is $U(0,1)$.

Let X have the pdf

$$f(x) = \frac{x^2}{3}, \quad -1 < x < 2.$$

Then the random variable $Y = X^2$ will have the support $0 \le y < 4$. Now, on the one hand, for $0 < y < 1$, we obtain the two-to-one transformation represented by $x_1 = -\sqrt{y}$ for $-1 < x_1 < 0$ and by $x_2 = \sqrt{y}$ for $0 < x_2 < 1$. On the other hand, if $1 < y < 4$, the one-to-one transformation is represented by $x_2 = \sqrt{y}, 1 < x_2 < 2$. Since

$$\frac{dx_1}{dy} = \frac{-1}{2\sqrt{y}} \quad \text{and} \quad \frac{dx_2}{dy} = \frac{1}{2\sqrt{y}},$$

it follows that the pdf of $Y = X^2$ is

$$g(y) = \begin{cases} \dfrac{(-\sqrt{y})^2}{3}\left|\dfrac{-1}{2\sqrt{y}}\right| + \dfrac{(\sqrt{y})^2}{3}\left|\dfrac{1}{2\sqrt{y}}\right| = \dfrac{\sqrt{y}}{3}, & 0 < y < 1, \\ \dfrac{(\sqrt{y})^2}{3}\left|\dfrac{1}{2\sqrt{y}}\right| = \dfrac{\sqrt{y}}{6}, & 1 < y < 4. \end{cases}$$

Note that if $0 < y < 1$, the pdf is the sum of two terms, but if $1 < y < 4$, then there is only one term. These different expressions in $g(y)$ correspond to the two types of transformations, namely, the two-to-one and the one-to-one transformations, respectively.