

**Inclusion-Exclusion Principle:**

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} A_{i_2}) + \dots + (-1)^{r+1} \sum_{1 \leq i_1 < \dots < i_r \leq n} P(A_{i_1} \dots A_{i_r}) + \dots + (-1)^{n+1} P(A_1 \dots A_n)$$

**General Multiplication Rule:**  $P(A_1 A_2 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \dots P(A_n|A_1 A_2 \dots A_{n-1})$

**Total Probability:**  $P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$

**Bayes' formula:** Events  $A_1, \dots, A_n$  partitions sample space, assume  $P(A_i) > 0$  for  $1 \leq i \leq n$ . Let  $B$  be any event, then for any  $1 \leq i \leq n$ , we have  $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n)}$

**Probability mass function:**  $p_X(x) = \begin{cases} P(X=x) & \text{if } x = x_1, x_2, \dots \\ 0 & \text{otherwise} \end{cases}$

**Cumulative distribution function:**  $F_X(x) = P(X \leq x)$  for  $x \in \mathbb{R}$

**Expected Value:**  $E(X) = \sum_x x p_X(x)$ ,  $E[g(x)] = \sum_i g(x_i) p_X(x_i) = \sum_x g(x) p_X(x)$

**Tail Sum Formula:**  $E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$

**Variance:**  $\text{var}(X) = E(X - \mu)^2 = E(X^2) - [E(X)]^2$

**Expected Value of Sum of RV:**  $E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i]$

**Bernoulli random variable:**  $Be(p)$ ,  $X = 1$  if success, 0 if failure.

$$P(X=1) = p, P(X=0) = 1-p, \quad \mathbb{E}(X) = p, \text{var}(X) = p(1-p)$$

**Binomial random variable:**  $Bin(n, p)$ ,  $X = \#$  of successes in  $n$  Bernoulli( $p$ ) trials.

$$\text{For } 0 \leq k \leq n, P(X=k) = \binom{n}{k} p^k q^{n-k} \quad \mathbb{E}(X) = np, \text{var}(X) = np(1-p)$$

**Geometric random variable:**  $Geom(p)$ ,  $X = \#$  of Bernoulli( $p$ ) trials required to obtain the first success.

$$\text{For } k \geq 1, P(X=k) = pq^{k-1} \quad \mathbb{E}(X) = \frac{1}{p}, \text{var}(X) = \frac{1-p}{p^2}.$$

**OR,**  $X' = \#$  of failures in Bernoulli( $p$ ) trials to obtain 1st success.  $X = X' + 1$

$$\text{For } k \geq 0, P(X'=k) = pq^k, \quad \mathbb{E}(X') = \frac{1-p}{p}, \text{var}(X') = \frac{1-p}{p^2}$$

**Negative Binomial random variable:**  $NB(r, p)$ ,  $X = \#$  of Bernoulli( $p$ ) trials required to obtain  $r$  success.

$$\text{For } k \geq r, P(X=k) = \binom{k-1}{r-1} p^r q^{k-r}, \quad \mathbb{E}(X) = \frac{r}{p}, \text{var}(X) = \frac{r(1-p)}{p^2}$$

$$\text{Note that } Geom(p) = NB(1, p), \quad \binom{k-1}{r-1} = (-1)^{r-1} \binom{-(k-r+1)}{r-1}$$

**Poisson Random Variable:**  $X \sim \text{Poisson}(\lambda)$  For  $k \geq 0$ ,  $P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}$ ,  $\mathbb{E}(X) = \lambda$ ,  $\text{var}(x) = \lambda$

Usually if  $n > 20$  and  $np < 15$ ,  $Bin(n, p) \approx \text{Poisson}(np)$ .

**Hypergeometric Random Variable:**  $H(n, N, m)$ , a set of  $N$  balls, of which  $m$  are red and  $N-m$  are blue. We choose  $n$  of these balls *without replacement*,  $X = \#$  of red balls in sample.

$$\text{For } 0 \leq x \leq \min(m, n), P(X=x) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}} \quad \mathbb{E}(X) = \frac{nm}{N}, \text{var}(X) = \frac{nm}{N} \left[ \frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N} \right]$$

**Expectation and Variance of Continuous RV:**

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf_X(x)dx, \quad \text{var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f_X(x)dx = \mathbb{E}[(x - \mu_X)^2]$$

$$\text{Tail sum formula: } \mathbb{E}(X) = \int_0^{\infty} P(X > x)dx = \int_0^{\infty} P(X \geq x)dx$$

$$\text{Uniform Distribution: } X \sim U(a, b), \quad f(x) = \frac{1}{b-a}, \quad a < x < b, \quad \mathbb{E}(X) = \frac{a+b}{2}, \quad \text{var}(x) = \frac{(b-a)^2}{12}$$

$$F_X(x) = 0, \text{ if } x < a; \quad \frac{x-a}{b-a}, \text{ if } a \leq x < b; \quad 1, \text{ if } b \leq x$$

$$\text{Normal distribution: } X \sim N(\mu, \sigma^2), \quad f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad \mathbb{E}(X) = \mu, \quad \text{var}(X) = \sigma^2$$

$$\text{Standard normal distribution: } X \sim N(0, 1), \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$$

$$P(a < Z < b) = \Phi(b) - \Phi(a), \quad P(Z < b) = \Phi(b), \quad \Phi(-x) = 1 - \Phi(x)$$

$$\frac{X - \mu}{\sigma} \sim N(0, 1), \quad \therefore Y \sim N(\mu, \sigma^2) \Rightarrow P(a < Y \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$\text{Exponential distribution: } X \sim \text{Exp}(\lambda), \quad f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \text{the c.d.f is } F_X(x) = 1 - e^{-\lambda x}, \quad x > 0$$

$$\mathbb{E}(X) = \frac{1}{\lambda}, \quad \text{var}(X) = \frac{1}{\lambda^2}$$

$$\text{memoryless property of exp dist: } P(X > s + t | X > s) = P(X > t), \quad s, t > 0$$

$$\text{Gamma distribution: } X \sim \Gamma(\alpha, \lambda), \quad f_X(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, \quad x \geq 0, \quad \text{where } \lambda, \alpha > 0, \quad \Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1} dy$$

$$\mathbb{E}(X) = \frac{\alpha}{\lambda}, \quad \text{var}(X) = \frac{\alpha}{\lambda^2}, \quad \Gamma(1) = 1, \quad \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1), \quad \Gamma(n) = (n-1)!, \quad \Gamma(1, \lambda) = \text{Exp}(\lambda), \quad \Gamma(1/2) = \sqrt{\pi}$$

$$\text{Beta distribution: } X \sim \text{Beta}(a, b), \quad f(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1, \quad \text{where beta function } B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx,$$

$$\mathbb{E}(X) = \frac{a}{a+b}, \quad \text{var}(X) = \frac{ab}{(a+b)^2(a+b+1)}, \quad B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\text{Cauchy distribution: } X \sim \text{Cauchy}(\theta), \quad f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty, \quad \mathbb{E}(X) = \infty, \quad \text{var}(X) = \infty$$

$$\text{De Moivre-Laplace Limit Thm: } X \sim \text{Bin}(n, p), \quad \text{then for any } a < b, \quad \text{Bin}(n, p) \approx N(np, npq)$$

$$P\left(a < \frac{X - np}{\sqrt{npq}} \leq b\right) \approx \Phi(b) - \Phi(a)$$

$$\text{Continuity Correction: } X \sim \text{Bin}(n, p), \quad Z \sim N(0, 1), \quad \text{then}$$

$$P(a \leq X \leq b) \approx P\left(\frac{a - 0.5 - np}{\sqrt{npq}} \leq Z \leq \frac{b + 0.5 - np}{\sqrt{npq}}\right), \quad P(a < X < b) \approx P\left(\frac{a + 0.5 - np}{\sqrt{npq}} \leq Z \leq \frac{b - 0.5 - np}{\sqrt{npq}}\right)$$

$$\text{Dist of a func of a RV: For monotonic } Y = g(X), \quad f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

$$\text{If } X \text{ is a RV with c.d.f } F, \quad \text{then } F(X) \sim U(0, 1).$$

$$\text{Marginal distribution function: } F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y) = \lim_{y \rightarrow \infty} P(X \leq x, Y \leq y) = P(X \leq x). \quad (\text{c.d.f of } X)$$

$$\text{Marginal p.m.f: } p_X(x) = P(X = x) = \sum_{y \in \mathbb{R}} p_{X,Y}(x, y) \quad \text{Marginal p.d.f: } f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

$$\text{Relation between p.d.f and c.d.f: } f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

$$\text{Independent: } p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

$$\text{Sum of Indep: } F_{X+Y}(x) = \int_{-\infty}^{\infty} F_X(x-t)f_Y(t)dt = \int_{-\infty}^{\infty} F_Y(x-t)f_X(t)dt,$$

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(x-t)f_Y(t)dt = \int_{-\infty}^{\infty} f_Y(x-t)f_X(t)dt$$

**Some conclusions:**  $X_1, \dots, X_n$  be  $n$  independent  $RV \sim \text{Exp}(\lambda)$ , then  $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$ .

$X_1, \dots, X_n$  be  $n$  independent  $RV \sim N(\mu_i, \sigma_i^2)$ , then  $\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$ . (used to approx Binominal Dist.)

**Sum of Discrete RV:**  $X \sim \text{Poisson}(\lambda), Y \sim \text{Poisson}(\mu)$ , then  $X + Y \sim \text{Poisson}(\lambda + \mu)$   
 $X \sim \text{Bin}(n, p), Y \sim \text{Bin}(m, p)$ , then  $X + Y \sim \text{Bin}(n + m, p)$   $X \sim \text{Geom}(p), Y \sim \text{Geom}(p)$ , then  $X + Y \sim NB(2, p)$

**Conditional Dist.:** (Discrete:)  $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$ ,  $F_{X|Y}(x|y) = P(X \leq x|Y = y)$

(Cont.):  $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ ,  $F_{X|Y}(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^x f_{X|Y}(t|y)dt$

**Joint p.d.f of Func of RV:**  $J(x, y) = \frac{\partial(u, v)}{\partial(x, y)} = \frac{\partial g}{\partial x} \frac{\partial h}{\partial y} - \frac{\partial g}{\partial y} \frac{\partial h}{\partial x}$ ,  $f_{U,V}(u, v) = f_{X,Y}(x, y)|J(x, y)|^{-1}$

**Expectation of Sum of RV:**  $\mathbb{E}[g(X, Y)] = \sum_y \sum_x g(x, y)p_{X,Y}(x, y)$ ,  $\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dxdy$

**Covariance:**  $\text{cov}(X, Y) = \mathbb{E}(X - \mu_X)(Y - \mu_Y)$ , if  $\text{cov}(X, Y) \neq 0$ , then  $X, Y$  are correlated.

$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ ,  $\text{cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{cov}(X_i, Y_j)$ ,

$\text{var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{var}(X_k) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j)$ , under **indep.**,  $\text{var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{var}(X_k)$

**Independent Case:**  $X, Y$  independent, then  $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$ ,  $\text{cov}(X, Y) = 0$ (reverse not true)

**correlation coefficient:**  $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$

**Conditional Expectation:**  $\mathbb{E}[X|Y = y] = \sum_x x p_{X|Y}(x|y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$ ,  $\mathbb{E}\left[\sum_{k=1}^n X_k|Y = y\right] = \sum_{k=1}^n \mathbb{E}[X_k|Y = y]$

**Expectation by Conditioning:**  $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}(X|Y)] = \sum_y \mathbb{E}(X|Y = y)P(Y = y) = \int_{-\infty}^{\infty} \mathbb{E}(X|Y = y)f_Y(y)dy$

**Probability by Conditioning:**  $P(A) = \sum_y P(A|Y = y)P(Y = y) = \int_{-\infty}^{\infty} P(A|Y = y)f_Y(y)dy$

**conditional variance:**  $\text{var}(X|Y) = \mathbb{E}[(X - \mathbb{E}[X|Y])^2|Y]$ ,  $\text{var}(X) = \mathbb{E}[\text{var}(X|Y)] + \text{var}(\mathbb{E}[X|Y])$   
 $\text{var}(Y|X) = \mathbb{E}(Y^2|X) - \mathbb{E}(Y|X)^2$

**Moment Generating Function:**  $M_X(t) = \mathbb{E}[e^{tX}] = \sum_X e^{tx} p_X(x) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$ ,  $\mathbb{E}(X^n) = M_X^{(n)}(0)$

If  $X, Y$  are independent,  $M_{X+Y}(t) = M_X(t)M_Y(t)$

**MGF for dist.:**  $X \sim \text{Be}(p)$ ,  $M(t) = 1 - p + pe^t$ ,  $X \sim \text{Bin}(n, p)$ ,  $M(t) = (1 - p + pe^t)^n$

$X \sim \text{Geom}(p)$ ,  $M(t) = \frac{pe^t}{1 - (1 - p)e^t}$ ,  $X \sim \text{Poisson}(\lambda)$ ,  $M(t) = \exp(\lambda(e^t - 1))$

$X \sim U(\alpha, \beta)$ ,  $M(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$ ,  $X \sim \text{Exp}(\lambda)$ ,  $M(t) = \frac{\lambda}{\lambda - t}$  for  $t < \lambda$ ,  $X \sim N(\mu, \sigma^2)$ ,  $M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

**Common Integrations:**  $\int x e^{ax} dx = \frac{1}{a^2}(ax - 1)e^{ax} + C$ ,  $\int x^n e^{ax} dx = \frac{1}{a}x^n e^{ax} - \frac{n}{a} \int x^{n-1} e^{ax} dx$