Rules of **Probability**

Independent

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

$$\begin{aligned} P(X \cap Y) = P(Y \cap X) &= P(X) \cdot P(Y \mid X) \\ &= P(Y) \cdot P(X \mid Y) \end{aligned}$$

Independent:

$$\begin{aligned} P(X \cap Y) = P(Y \cap X) &= P(X) \cdot P(Y) \\ &= P(Y) \cdot P(X) \end{aligned}$$

$$\begin{aligned} P(Y) &= 1 - P(Y') \\ &= P(Y \cap A) + P(Y \cap B) + ... + P(Y \cap Z) \\ &= P(Y \cap X) + P(Y \cap !X) \\ &= P(X) \cdot P(Y \mid X) + P(!X) \cdot P(Y \mid !X) \\ &= P(Y) \cdot P(X \mid Y) + P(Y) \cdot P(!X \mid Y) \end{aligned}$$

$$\begin{aligned} P(X \mid Y) &= \frac{P(X \cap Y)}{P(Y)} \\ &= \frac{P(X) \cdot P(Y \mid X)}{P(Y)} = \frac{P(Y) \cdot P(X \mid Y)}{P(Y)} \\ &= \frac{P(X) \cdot P(Y \mid X)}{P(X) \cdot P(Y \mid X) + P(!X) \cdot P(Y \mid !X)} = \frac{P(Y) \cdot P(X \mid Y)}{P(X) \cdot P(Y \mid X) + P(!X) \cdot P(Y \mid !X)} \\ &= \frac{P(X) \cdot P(Y \mid X)}{P(Y) \cdot P(X \mid Y) + P(Y) \cdot P(!X \mid Y)} = \frac{P(Y) \cdot P(X \mid Y)}{P(Y) \cdot P(X \mid Y) + P(Y) \cdot P(!X \mid Y)} \end{aligned}$$

Independent:

$$P(X \mid Y) = P(X)$$

$$\begin{aligned} P(X \cap Y \cap Z) &= P(X) \cdot P(Y \mid X) \cdot P(Z \mid X \cap Y) \\ &= P(Y) \cdot P(X \mid Y) \cdot P(Z \mid X \cap Y) \end{aligned}$$

$$P(X \cup Y \cup Z) = P(X) + P(Y) + P(Z) - P(X \cap Y) - P(X \cap Z) - P(Y \cap Z) + P(X \cap Y \cap Z)$$

$$P(A) + P(B) + P(C) = 1 \; if \; mutually \; exlusive$$

discrete

1. **probability mass function (PMF)**
   list all possible probabilities of x
   $p_X(x) = P(X = x)$
   (a) probability of x must be in [0, 1]
   (b) sum of all probabilities must equal to 1
   (c) R: hist(x)
   with x in x-axis and $p_X(x)$ in y-axis
2. **cumulative distribution function (CDF)**
   probability that a random variable is less than or equal to x
   $F_X(x) = P(X \leq x)$
   (a) $\lim_{x \to -\infty} F_X(x) = 0$
   (b) $\lim_{x \to \infty} F_X(x) = 1$
   (c) non-decreasing
3. $E_X(x)$ **= theoretical mean**
   For random variable x,
   $E_X(x) = \sum_{all\ x} x \cdot p_X(x)$
   For function g(x),
   $E_X(g(x)) = \sum_{all\ x} g(x) \cdot p_X(x)$

continuous

1. **probability density function (PDF)**
   list all possible probabilities of x
   $f_X(x) = \frac{\partial}{\partial x} F_X(x)$
   (a) $f_X(x) \geq 0$ for all x
   (b) area under the curve is equal to 1
   (c) R: plot(density(x))
   with x in x-axis and $f_X(x)$ in y-axis
2. **cumulative distribution function (CDF)**
   probability that a random variable is less than or equal to x
   $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)\ dx$
   $P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\ dx$
   (a) $\lim_{x \to -\infty} F_X(x) = 0$
   (b) $\lim_{x \to \infty} F_X(x) = 1$
   (c) non-decreasing
3. $E_X(x)$ **= theoretical mean**
   For random variable x,
   $E_X(x) = \int_{-\infty}^{\infty} x \cdot f_X(x)$
   For function g(x),
   $E_X(g(x)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)$

1. Rules of **Expectation**
   (a) $E(aX + b) = a \cdot E(X) + b$
   (b) $E(Y \mid X) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\ dy$
   (c) $E(XY) = E(E(XY \mid X)) = E(X \cdot E(Y \mid X))$
   (d) $E(XY) == E(X) \cdot E(Y)$ if X, Y independent
   (e) $E(XY \mid X) = X \cdot E(Y \mid X)$
   (f) $E(Y) = E(E(Y \mid X))$
   (g) $E_X(x^2) = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)$
2. Rules of **Variance**
   (a) $var(X) = \sigma^2 = E(X^2) - E(X)^2$
   (b) $var(X + c) = var(X)$
   (c) $var(cX) = c^2 \cdot var(X)$
   (d) $var(X - Y) = var(X) + var(Y)$
   (e) $var(X + Y) = var(X) + var(Y) + 2 \cdot cov(X, Y)$
   (f) $var(X + Y) = var(X) + var(Y)$ if X, Y independent
3. Rules of **Standard Deviation**
   (a) $\sigma = \sqrt{var(X)}$
4. Rules of **Normal Distribution**
   (a) $E(X) = 0$
   (b) $E(X^2) = 1$
   (c) $E(X)^2 = 0$
   (d) $\sigma = 1$
   (e) $var(X) = 1$
5. Rules of **Joint Distribution**
   Remember the rules of probability above (for discrete) applies here.
   (a) **Marginal**
   $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dy = \int_{-\infty}^{\infty} f_X(x) \cdot f_{Y|X}(y \mid x)\ dy$$
   $$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx = \int_{-\infty}^{\infty} f_Y(y) \cdot f_{X|Y}(x \mid y)\ dx$$
   (b) $f_{X|Y} = \frac{f_{X,Y}(X,Y)}{f_Y(y)}$
   (c) $f_{X,Y} == f_X(x) \cdot f_Y(y)$ if X, Y independent
6. Rules of **Covariance**
   Measure the strength of linear relationship b/w 2 variables
   (a) $cov(X, Y) = E((X - E(X)) \cdot (Y - E(Y)))$
   $= E(XY) - E(X) \cdot E(Y)$
   (b) $cov(X, Y) = 0$ if X, Y independent
   but cov == 0 does not mean X,Y independent
   (c) $cov(X, X) = var(X)$
   (d) $cov(aX, bY) = ab \cdot cov(XY)$
   (e) $cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$
7. Rules of **Correlation**
   Normalize Covariance by scaling it b/w -1 and 1
   (a) $corr(X, Y) = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}$
9. Rules of **Complement**
   (a) $P(X > 3) = 1 - P(X \le 3)$
   (b) $P(2 \le X \le 5) = P(X \le 5) - P(X \le 1)$
   (c) $P(2 < X < 5)$
   $= P(3 \le X \le 4) = P(X \le 4) - P(X \le 2)$
   (d) $P(A \mid B') = 1 - P(A \mid B)$
   (e) $P(A' \mid B') = 1 - P(A \mid B')$
10. Rules of **Sum**
   (a) Finite Arithmetic
   i.e. 3, 7, 11, ..., 51
   $S_n = \frac{n}{2}(a_1 + a_n) = n[a_1 + \frac{n-1}{2}d]$
   (b) Finite Geometric, $r = \frac{a_n}{a_n - 1}$
   i.e. 1, 2, 4, ..., 64
   $S_n = a_1 \frac{1 - r^n}{1 - r} = \frac{a_1 - r a_n}{1 - r}$
   (c) Infinite Geometric, $r = \frac{a_n}{a_n - 1}$, -1 < r < 1
   i.e. $3, 1, \frac{1}{3}, \frac{1}{9}, \frac{1}{27}, ...$
   $S = \frac{a_1}{1 - r}$
   (d) Infinite Arithmetic and Infinite Geometric w/ $r \le -1$ or $1 \le r$
   No formula
11. Rules of **Log**
   $$log_2(2^k) = log_2(9801)$$
   $$k = \frac{log_{10}(9801)}{log_{10}(2)}$$

Discrete Joint Probability Table

	y_1	y_2	...	y_n	
x_1	$f_{X,Y}(x_1, y_1)$	$f_{X,Y}(x_1, y_2)$	...	$f_{X,Y}(x_1, y_n)$	$f_X(x_1)$
x_2	$f_{X,Y}(x_2, y_1)$	$f_{X,Y}(x_2, y_2)$	...	$f_{X,Y}(x_2, y_n)$	$f_X(x_2)$
...	...	...	...	...	...
x_m	$f_{X,Y}(x_m, y_1)$	$f_{X,Y}(x_m, y_2)$	...	$f_{X,Y}(x_m, y_n)$	$f_X(x_m)$
	$f_Y(y_1)$	$f_Y(y_2)$	...	$f_Y(y_n)$	1

Variance-Covariance Matrix

$var(x_1)$	$cov(x_1, x_2)$	...	$cov(x_1, x_n)$
$cov(x_2, x_1)$	$var(x_2)$	...	$cov(x_2, x_n)$
...	...	...	...
$cov(x_n, x_1)$	$cov(x_n, x_2)$	...	$var(x_n)$

Distribution	Type	Description	PMF or PDF	$E_X(x)$	$var(x)$
Bernoulli	Discrete	Binomial with n = 1	$p^x(1-p)^{1-x}$	p	$1-p$
Binomial	Discrete	n independent Bernoulli trials	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$
Geometric	Discrete	get first success in independent Bernoulli trials	$(1-p)^{x-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	Discrete	get r^{th} success in independent Bernoulli trials	$\binom{x-1}{n-1}p^r(1-p)^{x-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometric	Discrete	sampling objects w/o replacement	$\frac{\binom{A}{a}\binom{B}{b}}{\binom{N}{n}} = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$	$\frac{n \cdot a}{N}$	$\frac{n \cdot a \cdot (N-a) \cdot (N-n)}{N^2 \cdot (N-1)}$
Poisson	Discrete	# of occurences of an event	$\frac{\lambda^x \cdot e^{-\lambda}}{x!}$	λ	λ
Uniform	Continuous	constant probability	$\frac{1}{b-a}, a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

Maximum Likelihood Estimator

1. p.d.f

   Poisson p.d.f

   $f(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$
2. likelihood function

$$\begin{aligned} L(\lambda) &= P(x_1 = x_1 \mid \lambda) \cdot ... \cdot P(x_n = x_n \mid \lambda) \\ &= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \end{aligned}$$

3. log of likelihood

$$\begin{aligned} l(\lambda) &= \ln(L(\lambda)) \\ &= \ln(\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}) \\ &= \sum_{i=1}^{n} x_i \cdot ln(\lambda) - \lambda - \ln(x_i!) \end{aligned}$$

4. λ_{MLE}

$$\begin{aligned} 0 &= \frac{\partial}{\partial \lambda} l(\lambda) \\ &= \frac{\partial}{\partial \lambda} \sum_{i=1}^{n} x_i \cdot ln(\lambda) - \lambda - \ln(x_i!) \\ &= \sum_{i=1}^{n} \frac{x_i}{\lambda} - 1 \\ &= (\sum_{i=1}^{n} \frac{x_i}{\lambda}) - n \\ &= \frac{1}{\lambda}(\sum_{i=1}^{n} x_i) - n \\ n &= \frac{1}{\lambda}(\sum_{i=1}^{n} x_i) \\ \lambda_{MLE} &= \frac{\sum_{i=1}^{n} x_i}{n} \end{aligned}$$

Z-test, T-test

Both tests measure the difference between an observed statistic and its hypothesized population parameter.

Z-value	T-value
measures in units of S.D.	measures in units of S.E.
knows population σ	only knows sample s

single value $z = \frac{x-\mu}{\sigma}$

mean of n obs $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ $\quad$ $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

$(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$ $\quad$ $(\bar{x} - t_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$