1 Distrbutions

Bernoulli: $\begin{cases} 1-p & \text{if } x=0 \\ p & \text{if } x=1 \end{cases}$

$$\mu = p, Var = pq$$

$$MGF = q + pe^t$$

Binomial: $\binom{n}{k}p^k q^{n-k}$

$$\mu = np, Var = npq$$

$$MGF = (q + pe^t)^n$$

Geometric: $p(1-p)^{k-1}$

$$\mu = \frac{1}{p}, Var = \frac{1-p}{p^2}$$

$$MGF = \frac{pe^t}{1-(1-p)e^t}$$

Poisson: $\frac{e^{-\lambda}\lambda^k}{k!}$

$$\mu = \lambda, Var = \lambda$$

$$MGF = e^{\lambda(e^t-1)}$$

Normal: $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(k-\mu)^2}{2\sigma^2}}$

$$\mu = \mu, Var = \sigma^2$$

$$MGF = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Exponential: $\lambda e^{-\lambda k}$

$$\mu = \frac{1}{\lambda}, Var = \frac{1}{\lambda^2}$$

$$MGF = \frac{\lambda}{\lambda - t},\ t < \lambda$$

Uniform: $\frac{1}{b-a}, \quad a \le k \le b$

$$\mu = \frac{a+b}{2}, Var = \frac{(b-a)^2}{12}$$

$$MGF = \frac{e^{bt}-e^{at}}{t(b-a)}$$

Beta: $\frac{k^{\alpha-1}(1-k)^{\beta-1}}{B(\alpha,\beta)}, \quad 0 \le k \le 1$

$$\mu = \frac{\alpha}{\alpha+\beta}, Var = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$MGF = {}_2F_1(\alpha, \alpha+\beta; \alpha+\beta+1; t)$$

$\beta(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
$\Gamma(\alpha) = (\alpha-1)!,\ \alpha \in \mathbb{N}$
Gamma: $\frac{\beta^\alpha k^{\alpha-1}e^{-\beta k}}{\Gamma(\alpha)}$

$$\mu = \frac{\alpha}{\beta}, Var = \frac{\alpha}{\beta^2}$$

$$MGF = \left(\frac{\beta}{\beta-t}\right)^\alpha,\ t < \beta$$

2 Formulas

2.1 Probability Formulas

$A \perp B \Leftrightarrow \mathbb{P}(A\|B) = \mathbb{P}(A) \& \mathbb{P}(B\|A) = \mathbb{P}(B) \Leftrightarrow \mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$
Union Bound: $\mathbb{P}(\cup_{i=1}^n A_i) \le \sum_{i=1}^n \mathbb{P}(A_i)$
Bayes' Rule: $\mathbb{P}(A\|B) = \frac{\mathbb{P}(B\|A)\cdot\mathbb{P}(A)}{\mathbb{P}(B)}$
Law of Total Probability: $\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A\|B_i)\cdot\mathbb{P}(B_i)$
Chain Rule: $\mathbb{P}(A_1\cap A_2\cap\ldots\cap A_n) = \mathbb{P}(A_1)\cdot\mathbb{P}(A_2\|A_1)\cdot\ldots\cdot\mathbb{P}(A_n\|A_1\cap\ldots\cap A_{n-1})$
Conditional Independence: $A \perp B\|C \Leftrightarrow \mathbb{P}(A\|B\cap C) = \mathbb{P}(A\|C)$

2.2 Expected Value and Variance

$\mathbb{E}[aX+b] = a\mathbb{E}[X]+b$
$\mathrm{Var}(X) = \mathbb{E}[X^2]-\mathbb{E}[X]^2 = \mathbb{E}((X-\mathbb{E}(X))^2)$
$\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)$ if $X \perp Y$
$\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$

2.3 Moments and MGF

$\mu_k = \mathbb{E}(X^k)$
$\overline{\mu_k} = \mathbb{E}((X-\mathbb{E}(X))^k)$
$M_X(t) = \mathbb{E}(e^{tX})$
$M_{X+Y}(t) = M_X(t)\cdot M_Y(t)$ if $X \perp Y$
$\mu_k = M_X^{(k)}(0)$
$\mathbb{E}[\mathbb{E}[X\|Y]] = \mathbb{E}[X]$

2.4 function of random variables

$\mathbb{E}[g(X)] = \sum_x g(x)\cdot\mathbb{P}(X=x)$
$f_Y(y) = f_X(g^{-1}(y))\cdot\left|\frac{d}{dy}g^{-1}(y)\right|$

2.5 Common Distributions

$F_{X,Y}(x,y) = \mathbb{P}(X\le x, Y\le y)$
$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\partial y}F_{X,Y}(x,y)$
$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
$X \perp Y \Leftrightarrow f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y)$
Jointly Gaussian:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 -2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) +\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right)$$

$f_{X\|Y}(x\|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
$\mathbb{E}[X\|Y=y] = \int_{-\infty}^{\infty} x\cdot f_{X\|Y}(x\|y)dx$

2.6 Normal Distribution

if $X \sim \mathcal{N}(\mu_X,\sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y,\sigma_Y^2)$, then $X+Y \sim \mathcal{N}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$
if X and Y are jointly Gaussian, then $X\|Y=y \sim \mathcal{N}(\mu_X+\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y), \sigma_X^2(1-\rho^2))$

2.7 Covaariance and Correlation

$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))] = \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)$
$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$
if two random variables are linearly independent then $\mathrm{Cov}(X,Y)=0$
$\mathrm{Cov}(aX+b, cY+d) = ac\mathrm{Cov}(X,Y)$ (independent of a,b,c,d)
$\mu_{i,j} = \mathbb{E}(X^iY^j)$
$\overline{\mu_{i,j}} = \mathbb{E}((X-\mathbb{E}(X))^i(Y-\mathbb{E}(Y))^j)$

2.8 CLT and LLN

Central Limit Theorem:
$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$ as $n\to\infty$

Law of Large Numbers:
$\overline{X} \to \mu$ as $n\to\infty$

2.9 Functions of multiple Random Variables

$$J = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

$$f_{Y_1,\ldots,Y_m}(y_1,\ldots,y_m) = f_{X_1,\ldots,X_n}(g_1^{-1}(y_1,\ldots,y_m) \ldots, g_m^{-1}(y_1,\ldots,y_m))\cdot det(J)$$

3 Probability Inequalities

3.1 Markov's Inequality

For any non-negative random variable X and any $a>0$:

$$\mathbb{P}(X\ge a) \le \frac{\mathbb{E}[X]}{a}$$

3.2 Chebyshev's Inequality

For any random variable X and any $a>0$:

$$\mathbb{P}(|X-\mathbb{E}[X]|\ge a) \le \frac{\mathrm{Var}(X)}{a^2}$$

3.3 Chernoff Bound

For any random variable X and any $t>0$:

$$\mathbb{P}(X\ge(1+\delta)\mathbb{E}[X]) \le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[X]}$$

$$\mathbb{P}(X\le(1-\delta)\mathbb{E}[X]) \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mathbb{E}[X]}$$

3.4 Jensen's Inequality

For any random variable X and any convex function f:

$$f(\mathbb{E}[X]) \le \mathbb{E}[f(X)]$$

3.5 Cauchy-Schwarz Inequality

given two vectors x and y:

$$\left(\sum_{i=1}^n x_iy_i\right)^2 \le \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right)$$

For any random variables X and Y:

$$\mathbb{E}[XY]^2 \le \mathbb{E}[X^2]\cdot\mathbb{E}[Y^2]$$

and the equality holds if and only if $X=aY$ for some $a\in\mathbb{R}$.

4 Sample Statistics

4.1 Properties:

Bias: $\mathbb{E}[\hat{\theta}]-\theta$
Variance: $\mathrm{Var}(\hat{\theta}) = \mathbb{E}[(\hat{\theta}-\mathbb{E}[\hat{\theta}])^2]$
Mean Squared Error: $\mathrm{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta}-\theta)^2] = \mathrm{Var}(\hat{\theta})+\mathrm{Bias}^2$
Consistency: $lim_{n\to\infty}\hat{\theta} = \theta$

4.2 Estimators:

Sample Mean: $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$
Sample Variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2$

4.3 Maximum Likelihood Estimation:

$\hat{\theta}_{\mathrm{MLE}} = \arg\max_\theta \prod_{i=1}^n f(X_i\|\theta) = \arg\max_\theta \sum_{i=1}^n \log f(X_i\|\theta)$
$\Rightarrow \frac{\partial\mathcal{L}(\theta)}{\partial\theta}$

4.4 MMSE:

linear MMSE: $h(a,b) = \mathbb{E}[(X-aY-b)^2]$
objective: minimize $h(a,b)$
solution: $a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} = \rho\frac{\sigma_X}{\sigma_Y}$, $b = \mathbb{E}[X]-a\mathbb{E}[Y]$

multiple MMSE:

$$\begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} \\ \sigma_{X_1X_2} & \sigma_{X_2}^2 \end{bmatrix}\begin{bmatrix} a_1^* \\ a_2^* \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(X_1,Y) \\ \mathrm{Cov}(X_2,Y) \end{bmatrix}$$

Covariance Matrix
$\hat{Y}(X_1,X_2) = a_1^*(X_1-\mathbb{E}[X_1]) + a_2^*(X_2-\mathbb{E}[X_2]) + \mathbb{E}[Y]$

4.5 Hypothesis Testing:

Null Hypothesis: $H_0 : \theta = \theta_0$
Alternative Hypothesis: $H_1 : \theta \neq \theta_0$
Type I Error: Reject H_0 when it is true
Type II Error: Accept H_0 when it is false
P_value: Probability of observing the data or more extreme data given H_0 is true

4.6 Z-Test:

$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$
$\Rightarrow$ p-value $= 2 \cdot (1 - \Phi(|Z|))$

Confidence Interval for Mean: $\overline{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

4.7 T-test

: Single Sample T-test:
$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$
$\Rightarrow$ p-value $= 2 \cdot (1 - t_{n-1}(|T|))$
2-Sample T-test:
$T = \frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$
$\Rightarrow$ p-value $= 2 \cdot (1 - t_{n_1+n_2-2}(|T|))$
2-Sample T-test (Unequal Variance):
$T = \frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$
$\Rightarrow$ p-value $= 2 \cdot (1 - t_\nu(|T|))$
where $\nu = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$

4.8 Chi-Square Test:

Goodness of Fit Test:
$\chi^2 = \sum_{i=1}^{k} \frac{(O_i-E_i)^2}{E_i}$
$\Rightarrow$ p-value $= 1 - \chi^2_{k-1}(\chi^2)$

4.9 Fisher's Exact Test:

H_0 : The two samples are from the same distribution
H_1 : The two samples are from different distributions
$P = \frac{\binom{n_1}{x}\binom{n_2}{y}}{\binom{n_1+n_2}{x+y}}$
p_value: getting the observed data or more extreme data

4.10 Permutation Test:

H_0 : The two samples are from the same distribution
H_1 : The two samples are from different distributions
consider an arbitrary test statistic T
permute the labels of the samples to get the distrbution of T under H_0
then calculate the p-value as the probability of observing the data or more extreme data under H_0

4.11 Beysian Inference:

MAP:
$\hat{\theta}_{\text{MAP}} = \arg\max_\theta f(\theta \| X) = \arg\max_\theta f(X \| \theta) \cdot f(\theta)$

4.12 Linear Regression:

Simple Linear Regression: $Y = \beta_0 + \beta_1 X + \epsilon$
where $\epsilon \sim \mathcal{N}(0, \sigma^2)$
β_1 and β_0 can be estimated using the Maximum Likelihood Estimation

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$
$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$