1 Probability

1.1 Sets

Disjoint Sets:

$$A \cap B = \varnothing$$

Distribution:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

DeMorgan's Laws:

$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}$$

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$$

Axioms of Probability:

$$P(A) \geq 0$$

$$P(S) = 1$$

$$P(A_1 \cup A_2 \cup ... \cup A_n) = \sum_{i-1}^{n} P(A_i)$$

1.2 Counting

$$P_k^n = \frac{n!}{(n-k)!}$$

$$C_k^n = \frac{n!}{k!(n-k)!}$$

$$\binom{n}{n_1 n_2 ... n_k} = \frac{n!}{n_1! n_2! ... n_k!}$$

1.3 Probability Laws

Independence if any of the following hold:

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$
$$P(A \cap B) = P(A)P(B)$$

Conditional Probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Multiplicative Law of Probability:

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(B)P(A|B) \\ &= P(A)P(B) \qquad \text{(if A, B independent)} \end{aligned}$$

Additive Law of Probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = P(A) + P(B) \qquad \text{(if A, B mutually exclusive)}$$

Complement Probabilities:

$$P(A) = 1 - P(\overline{A})$$

Law of Total Probability:

$$P(A) = \sum_{i=1}^{k} P(A \mid B_i) P(B_i)$$

Bayes' Rule:

$$P(B_j \mid A) = \frac{P(A \mid B) P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i) P(B_i)}$$

2 Discrete RV

2.1 General case

For any discrete probability function, the following must hold:

$$0 \leq p(x) \leq 1 \qquad \text{(for all } x\text{)}$$
$$\sum_x p(x) = 1$$

Probability function:

$$P(X = x) = p(x)$$

Expected value:

$$E[X] = \sum_x x\ p(x)$$

$$E[g(X)] = \sum_x g(x)\ p(x)$$

$$E[c] = c$$

$$E[aX + b] = aE[X] + b$$

Variance:

$$\begin{aligned} Var[X] &= E[(X - \mu)^2] \\ &= E[X^2] - E^2[X] \\ &= \sigma^2 \end{aligned}$$

$$Var[aX + b] = a^2\ Var[X]$$

2.2 Binomial Distribution

Binomial experiment criteria:

1. n identical trials.
2. Each trial either succeeds or fails.
3. Probability of each trial succeeding is p. Probability of failure is $q = 1 - p$.
4. Trials are independent.
5. Y denotes the number of successes observed during the n trials.

$$p(y) = \binom{n}{y} p^y q^{n-y}$$

$$E[Y] = np$$

$$Var[Y] = npq$$

2.3 Geometric Distribution

Geometric experiment criteria:

1. n identical trials.
2. Each trial either succeeds or fails.
3. Probability of each trial succeeding is p. Probability of failure is $q = 1 - p$.
4. Trials are independent.
5. Y denotes the number of trials up to the first success.

$$p(y) = q^{y-1}p$$
$$E[Y] = \frac{1}{p}$$
$$Var[Y] = \frac{1-p}{p^2}$$

2.4 Poisson Distribution

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda}$$
$$E[Y] = \lambda$$
$$Var[Y] = \lambda$$

3 Continuous RV

3.1 General case

CDF:

$$F(y) = P(Y \leq y)$$

PDF:

$$f(y) = \frac{d}{dy}F(y) = F'(y)$$
$$P(a \leq Y \leq b) = \int_a^b f(y)\, dy$$

Expected Value:

$$E[Y] = \int_{-\infty}^{\infty} y\, f(y)\, dy$$
$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)\, f(y)\, dy$$

3.2 Uniform Distribution

$$f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2 \\ 0, & \text{else} \end{cases}$$
$$E[Y] = \frac{\theta_1 + \theta_2}{2}$$
$$Var[Y] = \frac{(\theta_2 - \theta_1)^2}{12}$$

3.3 Normal Distribution

If $Y = Normal(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$
$$E[Y] = \mu$$
$$Var[Y] = \sigma^2$$

If $Y = Normal(0, 1)$:

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$$
$$E[Y] = 0$$
$$Var[Y] = 1$$

Normalize:

$$Z = \frac{X - \mu}{\sigma}$$

Phi function:

$$\Phi(y) = \text{The area under the normal curve to the left of } y$$

3.4 Exponential Distribution

$$f(x) = \lambda e^{-\lambda x}$$
$$F(x) = 1 - e^{-\lambda x}$$
$$E[X] = \frac{1}{\lambda}$$
$$Var[X] = \frac{1}{\lambda^2}$$

Memoryless property:

$$memoryless$$

4 Random Vectors

4.1 Discrete Bivariate Probability

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$$
$$p(y_1, y_2) \geq 0 \text{ for all } y_1, y_2$$
$$\sum_{y_1, y_2} p(y_1, y_2) = 1$$

4.2 Continuous Bivariate Probability

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2)$$
$$F(y_1, y_2) = \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2\, dt_1$$
$$F(-\infty, -\infty) = 0$$
$$F(\infty, \infty) = 1$$
$$f(y_1, y_2) \geq 0 \text{ for all } y_1, y_2$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1\, dy_2 = 1$$

4.3 Marginal Probability

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2)$$
$$p_2(y_2) = \sum_{y_1} p(y_1, y_2)$$
$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2$$
$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1$$

4.4 Independent Random Variables

The following are all true for independent random variables:

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)$$
$$p(y_1, y_2) = p_1(y_1)p_2(y_2)$$
$$f(y_1, y_2) = f_1(y_1)f_2(y_2)$$
$$f(y_1, y_2) = g(y_1)h(y_2)$$
$$Cov(Y_1, Y_2) = 0$$

4.5 Multivariate Expected Value

$$E[g(Y_1, ..., Y_k)] = \sum_{y_k} ... \sum_{y_1} g(y_1, ..., y_k)p(y_1, ..., y_k)$$
$$E[g(Y_1, ..., Y_k)] = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(y_1, ..., y_k)p(y_1, ..., y_k)\, dy_1 ... \, dy_k$$
$$E[c] = c$$
$$E[cg(Y_1, Y_2)] = cE[g(Y_1, Y_2)]$$
$$E[g_1(Y_1, Y_2) + ... + g_k(Y_1, Y_2)] = E[g_1(Y_1, Y_2)] + ... + E[g_k(Y_1, Y_2)]$$
$$E[g_1(Y_1, Y_2)g_2(Y_1, Y_2)] = E[g_1(Y_1, Y_2)]E[g_2(Y_1, Y_2)]$$

4.6 Covariance

$$Cov(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$
$$Cov(Y_1, Y_2) = E[Y_1Y_2] - E[Y_1]E[Y_2]$$
$$\rho = \frac{Cov(Y_1, Y_2)}{\sigma_1\sigma_2}$$

4.7 Conditional Probability

$$p(y_1 \mid y_2) = \frac{p(y_1, y_2)}{p_2(y_2)}$$
$$F(y_1 \mid y_2) = P(Y_1 \leq y_1, Y_2 = y_2)$$
$$f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$
$$E[X \mid Y] = \int_{-\infty}^{\infty} xf(x \mid y)\, dx$$
$$E[h(X) \mid Y] = \int_{-\infty}^{\infty} h(x)f(x \mid y)\, dx$$

Tower Property:

$$E[Y] = E[E[Y \mid X]]$$

4.8 Joint Normal Distribution

Definition:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} = N(\mu, \Sigma), \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1d} \\ \Sigma_{21} & \Sigma_{22} & & \vdots \\ \vdots & & & \vdots \\ \Sigma_{d1} & \cdots & \cdots & \Sigma_{dd} \end{bmatrix}$$

Properties:

$$E[X] = E\left[\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}\right] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_d] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}$$

$$\Sigma_{ij} = \Sigma_{ji} = Cov(X_i, X_j) = Cov(X_j, X_i)$$

Theorem: All linear transforms of a jointly normal distribution are also jointly normal.
Theorem: If X and Y are jointly normal, and $Cov(X, Y) = 0$, X and Y are independent.

5 Statistical Inference

5.1 Critical Theorems

Strong Law of Large Numbers:

$$\lim_{n \to \infty} \overline{X}_n = \mu \implies \overline{X}_n \approx \mu \qquad \text{(Big } n\text{)}$$

Central Limit Theorem:

$$\lim_{n \to \infty} \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = N(0, 1) \implies \overline{X}_n \approx N(\mu, \frac{\sigma^2}{n}) \qquad \text{(Big } n\text{)}$$

5.2 Estimation

$$\hat{\theta} = T(X_1, \ldots, X_n)$$

Consistency:

$$\lim_{n\to\infty} \hat{\theta} = \theta$$

Bias:

$$Bias[\hat{\theta}] = E[\hat{\theta}] - \theta$$

Mean Squared Error:

$$MSE[\hat{\theta}] = E[(\hat{\theta} - \theta)^2] = Bias^2[\hat{\theta}] - Var[\hat{\theta}]$$

Sample Mean:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{p}$$

Sample Variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Confidence Interval (Generic):

$$[L, R] = \hat{\theta} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \qquad \text{(Where P}(\theta \in [L, R] = 1 - \alpha))$$

Likelihood Function:

$$L_\theta = \text{Joint PDF/PMF for } (X_1, \ldots, X_n)$$

Statistic:

$$T = T(X_1, \ldots, X_n)$$

T is sufficient if:

$$L_\theta = g_\theta(X_1, \ldots, X_n) \cdot h(X_1, \ldots, X_n)$$

Minimum Variance Unbiased Estimate (MVUE):

$$\hat{\theta}^* = E[\hat{\theta} \mid T] = E[\hat{\theta}(X_1, \ldots, X_n) \mid T(X_1, \ldots, X_n)]$$

Rao-Blackwell Theorem:

$$E[\hat{\theta}^*] = \theta$$

$$V[\hat{\theta}^*] = V[\theta]$$

Maximum Likelihood Estimate (MLE):

$$\hat{\theta}_{MLE} = \theta \text{ that maximizes } L_\theta \text{ or } \ln(L_\theta)$$

Note: MLE is consistent under mild conditions and is asymptotically normal.

5.3 Hypothesis Testing

Null Hypothesis:

$$H_0$$

Alternative Hypothesis:

$$H_a$$

p-value:

Probability of observing something as or more extreme than the sample data, given H_0

Type 1 Error:

$$P(\text{Rejecting } H_0 \mid H_0 \text{ is true})$$

Type 2 Error:

$$P(\text{Accepting } H_0 \mid H_0 \text{ is false})$$