APMA 1655: Final Exam Review

Contents

1 Probability

1.1 Definitions

Random Event: an event with more than one possible outcome of varying likelihoods where the true outcome is a priori unknown

Sample Space Ω: the set of all possible outcomes of an experiment

Event: Each subset E of Ω

Impossible Event: the empty set $\emptyset$

1.2 Set Operations

Suppose Ω is a sample space and A and B are events $\{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$

Intersection $(A \cap B)$: A and B

Union $(A \cup B)$: A or B

Complement (A^c): not A

De Morgan's Laws:

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

Infinite Sets

Infinite Intersection: the collection of events that are in all the sets A_1, ..., A_n

$$\bigcap_{n=1}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n,\ \forall n = 1, 2, ...\}$$

Infinite union: the collection of events in at least one of the sets ("at least one of these events happens")

$$\bigcup_{n=1}^{\infty} A_n = \{\omega \in \Omega : \exists i\ | \omega \in A_i\}$$

1.3 Probability Space

Definitions

Disjoint: $A \cap B = \emptyset$

Mutually disjoint: all pairwise intersections of A_1, ..., A_n are empty: $A_n \cup A_m = \emptyset \quad n \neq m$

Probability $\mathbb{P}$: a real-valued function $\mathbb{P} : \{\text{subsets of } \Omega\} \to \mathbb{R}$. This is often defined as

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} \qquad A \subset \Omega$$

Probability space: The pair $(\Omega, \mathbb{P})$ if $\mathbb{P}$ satisfies the following three axioms:

1. $\mathbb{P}(A) \geq 0 \qquad \forall A \subset \Omega$
2. $\mathbb{P}(\Omega) = 1$
3. For any sequence of disjoint subsets $\{A_i\}_{i=1}^{\infty}$

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

1.4 Properties of Probability

1. $\mathbb{P}(\emptyset) = 0$
2. if $E_1 \cap E_2 = \emptyset$

$$\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$$

3. if $A, B \subset \Omega$ and $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. $0 \leq \mathbb{P}(A) \leq 1$
5. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
6. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
7.

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

1.5 Conditional Probability

If $\mathbb{P}(B) > 0$,

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

Theorem: If $(\Omega, \mathbb{P})$ is a probability space, $B \subset \Omega$, $\mathbb{P}(B) > 0$, then $(\mathbb{P}(A|B), \Omega)$ is also a probability space

Multiplication Law

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

Partition: $B_1, \ldots, B_n \subset \Omega$ if they are mutually disjoint and $\bigcup_{i=1}^n B_i = \Omega$

The Law of Total Probability: If $B_1, \ldots, B_n$ provide a partition of Ω

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$$

Corollary of the Law of Total Probability: If $0 < \mathbb{P}(B) < 1$,

$$\mathbb{P}(A) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) + \mathbb{P}(A|B^c) \cdot \mathbb{P}(B^c)$$

Bayes' Rule: Suppose $B_1, \ldots, B_n$ partition Ω and $\mathbb{P}(B_i), \mathbb{P}(A) > 0 \quad (i \in [1, n])$. Then

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)} \qquad i = 1, 2, \ldots, n$$

Independence: events that do not affect each other's outcomes:

$$\begin{cases} \mathbb{P}(A|B) = \mathbb{P}(A) \\ \mathbb{P}(B|A) = \mathbb{P}(B) \end{cases}$$

For independent events,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Mutually independent: if $\mathbb{P}(A_m \cap A_n) = \mathbb{P}(A_m) \cdot \mathbb{P}(A_n) \quad m \neq n$

1.6 Random Variable

Definition: On a probability space $(\Omega, \mathbb{P})$, a real valued function $X : \Omega \to \mathbb{R}$ is a random variable

Continuous Random Variable: a random variable with a continuous CDF

Discrete Random Variable: a random variable with a discrete CDF

Independent Random Vairbales: Y, Z on $(\Omega, \mathbb{P})$ are independent if

$$\mathbb{P}((Y \in A) \cap (Z \in B)) = \mathbb{P}(Y \in A) \cdot \mathbb{P}(Z \in B)$$

for any subsets $A, B \in \mathbb{R}$

1.7 Cumulative Distribution Functions (CDF)

$$F_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$$

Bernoulli Distribution: $X \sim \text{Bernoulli}(p)$ if

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Properties of the CDF F_X

1. $F_X(x_1) \leq F_X(x_2) \qquad x_1 \leq x_2$
2. $$\begin{cases} \lim_{x \to -\infty} F_X(x) = 0 \\ \lim_{x \to \infty} F_X(x) = 1 \end{cases}$$
3. F_X is right continuous $(F_X(x_0) = \lim_{x \to x_0^+} F_X(x))$
4. $\mathbb{P}(X = x_0) = F_X(x_0) - \lim_{x \to x_0^-} F_X(x)$ Note that this is zero if the CDF is continuous

Grand Theorem: Given a CDF, there exist a corresponding probability space and a random variable

1.8 Continuous Random Variable

For a continuous random variable and a real number x_0

$$\mathbb{P}(X = x_0) = 0$$

Theorem: if F_X is a CDF, it is piecewise differentiable

Probability density function (PDF): $p_X(x) = F'_X(x)$

For a continuous random variable X,

$$F_X(x) = \int_{-\infty}^{x} p_X(t)\ dt$$

Theorem: For X a continuous random variable with PDF p_X,

$$\int_{-\infty}^{\infty} p_X(t)\ dt = 1$$

Normal Distribution: $X \sim N(\mu, \sigma^2)$ (Normal distribution with mean μ and variance σ^2) if

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\ dt$$

1.9 Discrete Random Variables

For a discrete random variable X, its CDF is

$$F_X(x) = \sum_{k=1}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(x)$$

where $\{p_k\}_{k=1}^{K}$ is the **probability mass function** and is the probability $\mathbb{P}(X = x_k)$

Example: $X \sim \text{Bernoulli}(p)$

$$F_X(x) = (1-p) \cdot \mathbf{1}_{[0,\infty)}(x) + p \cdot \mathbf{1}_{[1,\infty)}$$

Poisson Distribution: $X \sim \text{Pois}(\lambda)$ if

$$F_X(x) = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \cdot \mathbf{1}_{[k,\infty)}(x)$$

1.10 Expected Value (Mean)

Discrete Version: IF X is a discrete RV and $\sum_{k=0}^{K} |x_t| \cdot p_k < \infty$ (i.e. if the sum if absolutely convergent), then

$$\mathbb{E}(X) = \sum_{k=0}^{K} x_k \cdot p_k$$

Continuous version: IF X is a continuous RV then if $\int_{-\infty}^{\infty} |x| \cdot p_X(x)\, dx < \infty$ then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx$$

For a permutation σ and an absolutely convergent series,

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} x_k \cdot p_k = \sum_{k=0}^{\infty} x_{\sigma(k)} \cdot p_{\sigma(k)}$$

In other words, the order of summation does not matter.

1.11 Transformations of RV

For a real-valued function $g(x)$, $g(X)$ is also a random variable.

Assuming the expected value exists,

- If X is discrete,
$$\mathbb{E}[g(X)] = \sum_{k=0}^{K} g(x_k) \cdot p_k$$
- If X is continuous,
$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot p_X(x)\, dx$$

Properties of Expected Values

- For a constant c, $\mathbb{E}(c) = c$
- For constants a, b,
$$\mathbb{E}[aX + b] = a \cdot \mathbb{E}(X) + b$$

- For $g_1(x)$, ..., $g_J(x)$ as functions where $\mathbb{E}[g_k(x)]$ exists for all $k = 1$, ..., J,

$$\mathbb{E}[g_1(X) + \ldots + g_J(x)] = \mathbb{E}\left[\sum_{k=1}^{J} g_k(x)\right] = \sum_{k=1}^{J} \mathbb{E}[g_k(X)]$$

  (the expected value is linear)

1.12 Variance

For a RV X that follows some distribution and generates numbers X_1, ..., X_n

Sample average:

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \approx \mathbb{E}X$$

Variance: For a RV X whose expected value exists,

$$\mathrm{Var}\, X = \mathbb{E}[(x - \mathbb{E}X)^2]$$

this can also be written

$$\mathrm{Var}\, X = \sum_{k=0}^{K} (x_k - \mathbb{E}X)^2 \cdot p_k \qquad \text{(X discrete)}$$

$$\mathrm{Var}\, X = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 \cdot p_X(x)\, dx \qquad \text{(X continuous)}$$

Properties of Variance

1. $\mathrm{Var}\, X = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
2. $\mathrm{Var}\,(aX + b) = a^2 \mathrm{Var}\,(X)$
3. For any constant c, $\mathrm{Var}\, c = 0$
4. If $\mathrm{Var}\, X = 0$ then there exists a c such that $\mathbb{P}(X = c) = 1$

1.13 The Law of Large Numbers (LLN)

Theorem: Let $\{X_i\}_{i=1}^{\infty}$ be an infinitely long sequence of independently and identically distributed RVs defined on $(\Omega, \mathbb{P})$, then

$$\mathbb{P}\left(\{\omega \in \Omega : \lim_{n\to\infty} \frac{X_1(\omega) + X_2(\omega) + \ ... + X_n(\omega)}{n} = \mathbb{E}X_1\}\right) = 1$$

where $\mathbb{E}X_1 = \mathbb{E}X_2 = ...$ because the CDFs are equal (by identical distribution)

Generalized Theorem: Let $\{X_i\}_{i=1}^{\infty}$ be iid RVs defined on $(\Omega, \mathbb{P})$. If $\mathbb{E}[g(X_1)]$ exists, then

$$\mathbb{P}\left(\{\omega \in \Omega : \lim_{n\to\infty} \frac{g(X_1(\omega)) + g(X_2(\omega)) + \ ... + g(X_n(\omega))}{n} = \mathbb{E}[g(X_1)]\}\right) = 1$$

1.14 Monte Carlo Integration

If we seek to solve a very hard integral, e.g.,

$$I = \int_0^1 \cos^{-1}\left(\frac{\cos(\frac{\pi}{2}x)}{1 + 2\cos(\frac{\pi}{2}x)}\right)\, dx$$

We let $U \sim \text{Unif}(0,1)$ whose PDF is $\mathbf{1}_{[0,1)}(x)$. Denote the integrand $g(x)$. Then

$$I = \mathbb{E}[g(U)]$$

We can generate $X_1(\omega)$, $X_2(\omega), ... \overset{iid}{\sim} \text{Unif}(0,1)$ and with enough random values

$$\overline{g(X_n)} = \frac{g(X_1) + g(X_2) + \ ... + g(X_n)}{n} \approx \mathbb{E}[g(X_1)] = \int_0^1 \cos^{-1}\left(\frac{\cos(\frac{\pi}{2}x)}{1 + 2\cos(\frac{\pi}{2}x)}\right)\, dx$$

For integrals with bounds (a,b) rather than $(0,1)$ we can use the same method but define a new random variable from $U \sim \text{Unif}(0,1)$ where

$$X = a + (b-a) \cdot U \sim \text{Unif}(a,b)$$

1.15 Law of the Iterated Logarithm

Error: $e_n(\omega) = \overline{g(X_n)} - \mathbb{E}[g(X_1)]$

Theorem: Let $X_1, X_2, ...$ be iid RVs on $(\Omega, \mathbb{P})$ with EX_1 and $\operatorname{Var} X_1$ existing. Then (heuristically)

$$\mathbb{P}(\{\omega \in \Omega : |e_n(\omega)| \leq \sqrt{\operatorname{Var} X_i \cdot \frac{2\log(\log n)}{n}}\}) \approx 1$$
$$\mathbb{P}(\{\omega \in \Omega : |e_n(\omega)| > \sqrt{\operatorname{Var} X_i \cdot \frac{2\log(\log n)}{n}}\}) \approx 0$$

1.16 Central Limit Theorem

Theorem: Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid RVs on $(\Omega, \mathbb{P})$. Suppose $\mathbb{E}X_i$ and $\operatorname{Var} X_i$ exist. Define a sequence of random variables $\{G_n\}_{n=1}^{\infty}$ so that

$$G_n(\omega) = \sqrt{n} \cdot e_n(\omega) = \sqrt{n} \cdot (\overline{X}_n(\omega) - \mathbb{E}X_1)$$

Then the CDf of G_n converges to the CDF of $N(0, \operatorname{Var} X_1)$ as $n \to \infty$:

$$\lim_{n\to\infty} \mathbb{P}(G_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi \cdot \operatorname{Var} X_1}} \cdot \exp\left(-\frac{t^2}{2\operatorname{Var} X_1}\right)\, dt$$

Corollary: Under the same conditions,

$$\frac{G_n(\omega)}{\sqrt{\operatorname{Var} X_1}} = \sqrt{n} \cdot \frac{\overline{X}_n(\omega) - \mathbb{E}X_1}{\sqrt{\operatorname{Var} X - 1}} \dot{\sim} N(0,1)$$

Proof of the CLT

Weak Convergence $\mathbf{G_n} \xrightarrow{\mathbf{w}} \mathbf{G}$: A sequence $G_1, ..., G_n$ of RV converge weakly to a continuous RV G if

$$\lim_{n\to\infty} G_{G_n}(x) = F_G(x)$$

Strong Convergence: if

$$\lim_{n\to\infty} G_n(\omega) = G(\omega) \quad \forall \omega \in \Omega$$

Moment Generating Functions: For a RV X,

$$M_X(t) = \mathbb{E}[e^{tX}]$$

is the moment generating function.

k-th Moment of X:

$$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}X^k$$

For a sequence of RVs G_1, ..., G_n with continuous RV G,

$$\lim_{n\to\infty} M_{G_n}(t) = M_G(t) \Longrightarrow G_n \xrightarrow{w} G$$

Then if $G \sim N(0, \operatorname{Var} X_1)$, proof of the CLT simply depends on proving the convergence of the moment-generating functions.

Some Lemmas:

1. For $S_n(\omega) = \sum_{i=1}^n X_i(\omega)$ with X_i RVs, if X_1, ..., X_n are independent

$$M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t)$$

2. If the same RVs X_1, ..., X_n are also identically distributed

$$M_{S_n}(t) = (M_{X_1}(t))^n$$

3. For $\{C_n\}_{n=1}^\infty$ being a sequence of real-numbers for which $\lim_{n\to\infty} C_n = 0$, if $\lim_{n\to\infty} n \cdot C_n = \lambda$, then

$$\lim_{n\to\infty} (1 + C_n)^n = e^\lambda$$

4. If $G \sim N(0, \sigma^2)$ then

$$M_G(t) = \exp(\frac{t^2\sigma^2}{2})$$

Finally the proof:

$$\begin{aligned}
G_n &= \sqrt{n}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mathbb{E}X_1\right] \\
&= \sqrt{n}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \frac{1}{n}\cdot n\cdot \mathbb{E}X_1\right] \\
&= \frac{\sqrt{n}}{n}\left[\sum_{i=1}^{n} X_i - \sum_{i=1}^{n}\mathbb{E}X_1\right] \\
&= \frac{\sqrt{n}}{n}\left[\sum_{i=1}^{n}(X_i - \mathbb{E}X_1)\right]
\end{aligned}$$

Then

$$\begin{aligned}
M_{G_n}(t) = \mathbb{E}[e^{tG_n}] &= \mathbb{E}[\exp(\frac{t}{\sqrt{n}})\sum_{i=1}^{n}(X_i - \mathbb{E}X_1)] \\
&= \mathbb{E}\left[\exp\left(\sum_{i=1}^{n}\frac{t}{\sqrt{n}}(X_i - \mathbb{E}X_1)\right)\right] \\
&= \mathbb{E}\left[prod_{i=1}^{n}\exp\left(\sum_{i=1}^{n}\frac{t}{\sqrt{n}}(X_i - \mathbb{E}X_1)\right)\right] \quad \text{(by iid)} \\
&= \left(\mathbb{E}\left[\exp\left(\sum_{i=1}^{n}\frac{t}{\sqrt{n}}(X_i - \mathbb{E}X_1)\right)\right]\right)^n \\
&= \left(\mathbb{E}\left[1 + \frac{t}{\sqrt{n}}(X_1 - \mathbb{E}X_1) + \frac{t^2}{2n}(X_1 - \mathbb{E}X_1)^2 + \sum_{k=3}^{\infty}\frac{t^k}{k!n^{k/2}}(X_1 - \mathbb{E}X_1)^k\right]\right)^n \\
&= \left(1 + \underbrace{\frac{t^2}{2n}\operatorname{Var}X_1 + \sum_{k=3}^{\infty}\frac{t^k}{k!n^{k/2}}(X_1 - \mathbb{E}X_1)^k}_{C_n}\right)^n \quad \text{(because } \mathbb{E}[X_1 - \mathbb{E}X_1] = 0\text{)}
\end{aligned}$$

Using the lemmas above $\lim_{n\to\infty} C_n = 0$ and

$$n\cdot C_n = \lambda = \frac{t^2}{2}\operatorname{Var}X_1 + \sum_{k=3}^{\infty}\frac{t^k}{k!n^{\frac{k}{2}-1}}\mathbb{E}[(...)^k]$$

But when $k \geq 3$, $(\frac{k}{2} - 1) > 0$ so

$$\lim_{n\to\infty} n \cdot C_n = \frac{t^2}{\operatorname{Var} X_1} := \lambda$$

Then again by the lemmas

$$M_{G_n}(t) = (1 + C_n)^n \overset{n\to\infty}{\longrightarrow} e^\lambda = \exp(\frac{t^2}{2}\operatorname{Var} X_1)$$

But from the final lemma, the MGF of $N(0, \operatorname{Var} X_1)$ is

$$M_G(t) = \exp(\frac{t^2}{2}\operatorname{Var} X_1)$$

Thus,

$$M_{G_n}(t) \overset{n\to\infty}{M}_G(t)$$

and

$$G_n \overset{w}{\to} G \sim N(0, \operatorname{Var} X_1) \quad \blacksquare$$

1.17 Error Bounds

Let $\{X_i\}_{i=1}^\infty$ be a sequence of iid RVs on $(\Omega, \mathbb{P})$. Suppose $\mathbb{E}X_i$ and $\operatorname{Var} X_i$ exist.

From the law of the iterated logarithm, $|e_n(\omega) \leq \sqrt{2\log(\log n)} \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}}$ with probability around 100%.

$$\begin{aligned}
&\mathbb{P}\left(|e_n(\omega)| \leq z \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}}\right) \\
&= \mathbb{P}(-z \leq \sqrt{n} \cdot \frac{e_n(\omega)}{\sqrt{\operatorname{Var} X_1}} \leq z) \\
&= \mathbb{P}\left(\sqrt{n} \cdot \frac{e_n(\omega)}{\sqrt{\operatorname{Var} X_1}} \leq z\right) - \mathbb{P}\left(\sqrt{n} \cdot \frac{e_n(\omega)}{\sqrt{\operatorname{Var} X_1}} \leq -z\right) \\
&\approx \Phi(z) - \Phi(-z) \qquad (CLT) \\
&= 2\Phi(x) - 1
\end{aligned}$$

Where Φ is the CDF of $N(0,1)$.

Now let z^* denote the positive real number such that $\Phi(z^*) = 0.975$ so

$$\mathbb{P}\left(|e_n(\omega) \leq z^* \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}}\right) \approx 2\Phi(z^*) - 1 = 0.95$$

Generally, you can choose z^* such that $\Phi(z^*) = 1 - \frac{\alpha}{2}$ so $2\Phi(z^*) - 1 = 1 - \alpha$. Then z^* os the "$1 - \alpha/2$ quantile of $N(0, 1)$."

All together, this gives

$$|e_n(\omega)| \leq z^* \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}} \approx 1.96 \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}}$$

Conclusion: using the CLT we can establish much tighter error bounds than the LIL approach at the cost of only 5% confidence.

1.18 Random Vectors

Random Vector: a column vector $\vec{X} = (X_1, X_2 ..., X_n)^T$ defined on $(\Omega, \mathbb{P})$ if each of its components is a RV.

CDF of a Random Vector: an n-variable function

$$F_{\vec{X}}(x_1, x_2, \, ..., x_n) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{\omega \in \Omega : X_i(\omega) \leq x_i\}\right)$$

Continuous Random Vector: a random vector $\vec{X}$ if $F_{\vec{X}}$ is differentiable

The PDF of a Random Vector:

$$p_{\vec{X}}(x_1, \, ..., x_n) = \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} \cdots \frac{\partial}{\partial x_n} F_{\vec{X}}(x_1, x_2, \, ..., x_n)$$

Expected value of a Random vector: If $\vec{X} = (X_1, \, ..., X_n)^T$ is a continuous random vector with PDF $p_{\vec{X}}$, $g(\vec{x})$ is an n-variable function, $\int_{\mathbb{R}^n} |g(x_1 ..., x_n)| \cdot p_{\vec{X}}(x_1 ..., x_n) \, dx_1 ... dx_n < \infty$ then

$$\mathbb{E}[g(\vec{X})] = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ integrals}} g(x_1 ..., x_n) \cdot p_{\vec{X}}(x_1, \, ..., x_n) \, dx_1 ... dx_n$$

2 Statistics

2.1 Statistical Models

Sample Data: a collection $\{x_i\}_{i=1}^n = X_i(\omega^*)$ of deterministic numbers for some fixed $\omega^* \in \Omega$

Sample size: n in the definition of data

$\mathfrak{F}$-based model: Let $\mathfrak{F} = \{F_\theta\}_{\theta \in \Theta}$ be some family of real-valued functions satisfying the CDF properties. Then the $\mathfrak{F}$-based model is the assumption that there exists some "true" $\theta^* \in \Theta$ for which

$$X_1, X_2, \, ..., X_n \overset{iid}{\sim} F_{\theta^*}(x) = \mathbb{P}(X_1 \leq x)$$

Parameter Space: $\{F_\theta\}_{\theta \in \Theta}$, the family of functions from which we are selecting

Parametric Model: a model for which Θ is a subset of a finite-dimensional space

Non-parametric Model: a model for which Θ is a subset of an infinite-dimensional space

Unspecified model: a model for which the underlying assumption is incorrect (i.e. no true θ^* exists)

Statistical inference: the process of combining probability theory and data to infer the value of θ^*

2.2 Hypothesis Testing

We assume the $\{F_\theta\}_{\theta \in \Theta}$-based model is correct. We then let $\Theta = \Theta_0 \cup \Theta_1$, giving us two hypthoses (Θ_0 and Θ_1 partition Θ). Either:

1. **The Null Hypothesis:** $H_0 : \theta^* \in \Theta_0$
2. **The Alternative Hypothesis:** $H_1 : \theta^* \in \Theta_1$

Test: For sample size n, a test is any function $T : \mathbb{R}^n \to \{0, 1\}$. If $T(\vec{x}) = 1$ we reject H_0. If it is 0, we accept H_0. As T outputs in $\{0, 1\}$,

$$T(X_1(\omega), \, ..., X_n(\omega)) = R(\omega) \sim \text{Bernoulli}(r)$$

where $r = \mathbb{P}(R = 1) = \mathbb{E}R$

Type 1 Error: the null hypothesis is true ($\theta^* \in \Theta_0$) but we reject it ($T = 1$)

Type 2 Error: the null hypothesis is false ($\theta^* \in \Theta_1$) but we fail to reject it ($T = 0$)

Criteria for a Good Test

We define a function

$$\beta_T(\theta) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{n} T(\xi_1, \, ..., \xi_n) \cdot \prod_{i=1}^{n} p(\xi_i|\theta) \; d\xi_1 \dots d\xi_n$$

where $p(\xi_i|\theta) = F'_\theta(\xi_i)$ so $\beta_T(\theta^*) = \mathbb{E}[T(\vec{X})]$

This function can be interpreted "if F_θ is the true CDF, the probability of rejecting H_0 through T is $\beta_T(\theta)$"

Minimize Type 1 Error: make $\sup_{\theta \in \Theta}$ ("the significance of T") small

Minimize Type 2 Error: make $\beta_T(\theta)$ large for every $\theta \in \Theta_1$

2.3 Uniformly Most Powerful Test (UMP Test)

Definition: Let $\alpha \in (0, 1)$ be pre-specified. Suppose T^* is a test with significance α ($\sup_{\theta \in \Theta_0} \beta_{T^*}(\theta) = \alpha$). Then T^* is said to be a UMP test with significance α if for all T for which $\sup_{\theta \in \Theta_0} \beta_T(\theta) = \alpha$

$$\beta_T(\theta) \leq \beta_{T^*}(\theta) \quad \forall \theta \in \Theta_1$$

Neyman-Pearson Lemma: With $\Theta = \{\theta_0 \, , \theta_1\}, \;\; \Theta_0 = \{\theta_0\}, \;\; \Theta_1 = \{\theta_1\}$. Let $p(\xi|\theta) = F'_\theta(\xi)$ for all $\theta \in \Theta$. For any $\alpha \in (0, 1)$, the UMP test with significance alpha is

$$T_\alpha(\xi_1, \, ..., \xi_n) = \mathbf{1}\left(\frac{\prod_{i=1}^n p(\xi_i|\theta_1)}{\prod_{i=1}^n p(\xi_i|\theta_0)} > C_\alpha \right)$$

where C_α is the solution to $\beta_{T_{NP, \, \alpha}}(\theta_0) = \alpha$

2.4 The Maximum Likelihood Estimator

Point Estimating: the process of estimating θ^* in the $\mathfrak{F}$-based model, usually via MLE, the method of moments, or mean squared estimation.

The Likelihood Function: We assume the $\{F_\theta\}_{\theta\in\Theta}$-based model is correct. We assume that F_θ is piecewise differentiable for all θ and we have a collection of given, fixed, deterministic data $D = \{x_i\}_{i=1}^n$. Then

$$L(\theta|D) = \prod_{i=1}^{n} p(x_i|\theta)$$

The MLE: We select the θ which maximized the likelihood function

$$\widehat{\theta}_{MLE} = \arg\max_{\theta\in\Theta} L(\theta|D)$$

which by the consistency property is approximately equal to θ^*

Loss function: $-L(\theta|D)$ as

$$\arg\max_{\theta\in\Theta} L(\theta|D) = \arg\min_{\theta\in\Theta}\ -L(\theta|D)$$

Log-likelihood function:

$$l(\theta|D) = \log L(\theta|D) = \sum_{i=1}^{n} \log p(x_i|\theta)$$

Calculating the MLE

Example: $\Theta = \mathbb{R}$ and F_θ is the CDF of $N(0,1)$.

$$\begin{aligned}
p(\xi|\theta) &= \frac{1}{\sqrt{2\pi}}\exp(-\frac{(\xi-\theta)^2}{2}) \\
l(\theta|D) &= \sum_{i=1}^{n} \log p(x_i|\theta) \\
&= \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_i-\theta)^2}{2})\right) \\
&= \sum_{i=1}^{n}\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}(x_i-\theta)^2\right) \\
&= -\frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\frac{1}{2}(x_i-\theta)^2
\end{aligned}$$

The max will occur at the critical point so

$$\begin{aligned}
\widehat{\theta} &= \frac{\partial}{\partial\theta} l(\theta|D) = 0 \\
&= \frac{\partial}{\partial\theta}\left(-\frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\frac{1}{2}(x_i-\theta)^2\right) \qquad\qquad = \sum_{i=1}^{N}(x_i-\theta) \\
&= -n\cdot\theta + \sum_{i=1}^{n} x_i = 0 \\
&\implies \theta = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n
\end{aligned}$$

From the second derivative

$$\frac{\partial^2}{\partial\theta^2} l(\theta|D) = -n < 0$$

so this is the maximum

$$\widehat{\theta}_{MLE} = \overline{x}_n$$

Consistency of the MLE: Supposing that on $(\Omega, \mathbb{P})$, the $\{F_\theta\}_{\theta\in\Theta}$-based model is correct and the "regularity conditions" apply,

$$\forall\varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(\{\omega\in\Omega : |\widehat{\theta}(\omega) - \theta^*| < \varepsilon\}) = 1$$

where $\widehat{\theta}_n(\omega) = \widehat{\theta}_{MLE}(X_1(\omega), ..., X_n(\omega))$

Asymptotic normality of the MLE: under the same conditions as consistency,

$$\sqrt{n}(\widehat{\theta}_n - \theta^*)\dot{\sim}N(0, \frac{1}{I(\theta^*)})$$

where the "Fisher Information" $I(\theta)$ is

$$I(\theta) = \int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log p(\xi|\theta)^2\right)^2 \cdot p(\xi|\theta)\, d\xi$$