Formal review of Statistics

Moshi Alam

Three core sections

- Basic probability theory
- Expectations, Variance and covariance
- Hypothesis testing/Inference

Basic probability theory

Basic probability theory

Objects. Random process; sample space Ω; event $A \subseteq \Omega$. Discrete vs. continuous.

Independence. A and B are independent iff $\Pr(A \mid B) = \Pr(A)$.

Product rule (indep.). $\Pr(A, B) = \Pr(A) \Pr(B)$.

Examples.

- Cards (no replacement): $\Pr(\text{Ace on 2nd} \mid \text{Ace on 1st}) = \frac{3}{51}$.
- Dice (two fair dice): $\Pr(\text{sum} = 7) = \frac{6}{36}$, $\Pr(\text{sum} = 3) = \frac{2}{36}$.

Events and conditional probability

Set relations. Complement $\sim A$, union $A \cup B$, intersection $A \cap B$.

Conditional probability.

$$\Pr(B \mid A) = \frac{\Pr(A, B)}{\Pr(A)}, \qquad \Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)}.$$

Joint via conditional.

$$\Pr(A, B) = \Pr(A \mid B)\Pr(B) = \Pr(B \mid A)\Pr(A).$$

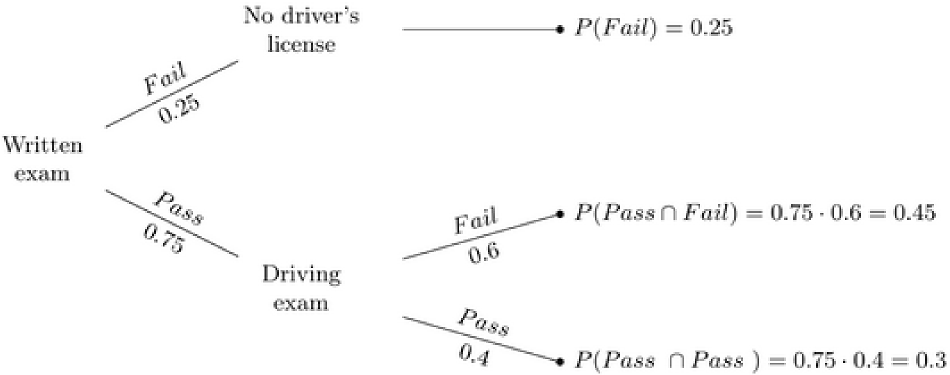
Law of total probability. For a partition $\{B_n\}$,

$$\Pr(A) = \sum_n \Pr(A \cap B_n) = \sum_n \Pr(A \mid B_n)\Pr(B_n).$$

Bayes (decomposition).

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \Pr(B \mid \sim A)\Pr(\sim A)}.$$

Probability Tree



Probability tree

Tree representation. Nodes: states; edges: conditional probabilities.

Reading a tree.

- Path probability = product of edge probabilities along the path.
- Sum of disjoint path probabilities = probability of the union.

Implication. Visual proof of law of total probability:

$$\Pr(A) = \sum_{b \in \mathcal{B}} \Pr(A \cap b) = \sum_{b} \Pr(A \mid b) \Pr(b).$$

Probability Trees

Driver's License Example:

- Must pass written exam first ($\Pr = 0.75$)
- Then take driving exam (if passed written)
- Joint probabilities sum to 1.0

Law of Total Probability:

$$\Pr(A) = \sum_n \Pr(A \cap B_n)$$

Conditional probability from tree:

$$\Pr(\text{Fail} \mid \text{Pass}) = \frac{0.45}{0.75} = 0.6$$

Set Theory and Venn Diagrams

Key Definitions:

- U: Universal set
- $A + \sim A = U$ (complement)
- $A \cup B$: Union (either A or B)
- $A \cap B$: Intersection (both A and B)

Set Relationships:

$$A = A \cap B + A \cap \sim B$$
$$A \cup B = A \cap \sim B + \sim A \cap B + A \cap B$$

Conditional Probability from Sets

Texas Football Coach Example:

- A: Team makes bowl game, $\Pr(A) = 0.6$
- B: Coach rehired, $\Pr(B) = 0.8$
- $\Pr(A, B) = 0.5$

Calculations:

$$\Pr(A, \sim B) = \Pr(A) - \Pr(A, B) = 0.6 - 0.5 = 0.1$$

$$\Pr(B \mid A) = \frac{\Pr(A, B)}{\Pr(A)} = \frac{0.5}{0.6} = 0.83$$

$$\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)} = \frac{0.5}{0.8} = 0.63$$

Contingency Tables

Two-way table structure:

Event	Not rehired ($\sim B$)	Rehired (B)	Total
Bowl game (A)	$\Pr(A, \sim B) = 0.1$	$\Pr(A, B) = 0.5$	$\Pr(A) = 0.6$
No bowl ($\sim A$)	$\Pr(\sim A, \sim B) = 0.1$	$\Pr(\sim A, B) = 0.3$	$\Pr(\sim A) = 0.4$
Total	$\Pr(\sim B) = 0.2$	$\Pr(B) = 0.8$	1.0

Joint Probability Definition:

$$\Pr(A, B) = \Pr(A \mid B) \Pr(B)$$
$$\Pr(B, A) = \Pr(B \mid A) \Pr(A)$$

Bayes's Rule

Naive Version:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

Bayesian Decomposition:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \Pr(B \mid \sim A)\Pr(\sim A)}$$

Updates prior beliefs with new information

- Prior: $\Pr(A)$
- Posterior: $\Pr(A \mid B)$

Monty Hall Setup

Problem:

- Three doors: D_1, D_2, D_3
- One has $1 million, two have goats
- You choose door 1
- Monty opens door 2 (reveals goat)
- Should you switch to door 3?

Intuition: Most people say "no difference" (50-50 chance)
Reality: You should switch!

Monty Hall Solution

Prior probabilities: $\Pr(A_i) = \frac{1}{3}$ for each door
Monty never opens door with money

$$\Pr(B \mid A_1) = 0.5 \text{ (could open door 2 or 3)}$$
$$\Pr(B \mid A_2) = 0.0 \text{ (never opens door with money)}$$
$$\Pr(B \mid A_3) = 1.0 \text{ (must open door 2)}$$

Results:

$$\Pr(A_1 \mid B) = \frac{1}{3} \text{ (your original choice)}$$
$$\Pr(A_3 \mid B) = \frac{2}{3} \text{ (switch choice)}$$

Switch and double your chances!

Expectation, Variances and Covariance

Notation

In this course for the most part we will be using a *single cross-sectional data*

1. Observations and Indices:
   - i: Index for a unit of observation ($i = 1, \ldots, n$)
     - Could be individual, city, firm
   - n: Sample size
   - $\sum_{i=1}^{n}$: Sum over all observations
2. Variables:
   - y: Outcome / dependent variable (what we want to explain)
   - x: Independent / explanatory variable(s) (what helps explain y)
   - y_i: The i-th observation of y
   - x_i: The i-th observation of x

Summation operator

Notation. $\sum_{i=1}^{n} x_i = x_1 + \cdots + x_n$.

Rules.

$$\sum_{i=1}^{n} c = nc, \qquad \sum_{i=1}^{n} c\,x_i = c\sum_{i=1}^{n} x_i, \qquad \sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i.$$

Summation Properties (cont.)

What summation is NOT:

$$\sum_{i=1}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$$

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

Useful results:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} (x_i - \overline{x}) = 0$$

$$\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i (y_i - \overline{y}) = \sum_{i=1}^{n} y_i (x_i - \overline{x})$$

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\overline{x})^2$$

Expected Value Definition

Expected value (discrete X) / Population mean:

$$E(X) = \sum_{j=1}^{k} x_j f(x_j)$$

where $f(x_j)$ is the probability of outcome x_j.

Example: $X \in \{-1, 0, 2\}$ with probabilities $\{0.3, 0.3, 0.4\}$

$$E(X) = (-1)(0.3) + (0)(0.3) + (2)(0.4) = 0.5$$
$$E(X^2) = (1)(0.3) + (0)(0.3) + (4)(0.4) = 1.9$$

Properties of Expected Value

Key properties:

1. $\mathbb{E}(c) = c$ for any constant c
2. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ for constants a, b
3. $\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i)$

Additional properties:

$$\mathbb{E}(W + H) = \mathbb{E}(W) + \mathbb{E}(H)$$

$$\mathbb{E}\big(W - \mathbb{E}(W)\big) = 0$$

Variance Definition

Population variance:

$$\mathrm{Var}(W) = \mathbb{E}\big[(W - \mathbb{E}[W])^2\big] = \mathbb{E}[W^2] - \mathbb{E}[W]^2.$$

Alternative formula:

$$V(W) = E(W^2) - [E(W)]^2$$

Sample variance (with degrees of freedom correction):

$$\widehat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Properties of Variance

Properties.

1. $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$
2. $\mathrm{Var}(c) = 0$, for any constant c
3. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.

Covariance

Definition.

$$\mathrm{Cov}(X,Y) = \mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Sign. $\mathrm{Cov}(X,Y) > 0$ (move together), < 0 (move opposite). Independence $\Rightarrow \mathrm{Cov}(X,Y) = 0.$

Algebra.

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y),$$

$$\mathrm{Cov}(a_1 + b_1 X,\ a_2 + b_2 Y) = b_1 b_2\ \mathrm{Cov}(X,Y).$$

Correlation.

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \in [-1,1].$$

Inference

Normal Distributions

Random variable: $X \sim N(\mu, \sigma^2)$
PDF: $f(X = x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, where

- $E(X) = \mu$ is the mean and Var$(X) = \sigma^2$ is the variance
- The distribution is symmetric around μ

Properties:

- $Z = \frac{X-\mu}{\sigma}$ is the standard normal random variable with mean 0 and variance 1
- Transformation of X such as $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- If X & Y are independent normal random variables, then $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
- Any linear combination of normal random variables is normal

χ^2_n and t_n

$Z_i \sim N(0,1)$ for $i = 1, 2, \ldots, n$ independent standard normal random variables

- The sum of squared independent standard normal random variables follows a chi-square distribution with n d.o.f. Let $Q = \sum_{i=1}^{n} Z_i^2$

$$Q \sim \chi^2_n$$

  - $E(Q) = n$ and Var$(Q) = 2n$
- The ratio of a standard normal random variable and a chi-square random variable follows a t-distribution. Let $T = \frac{Z}{\sqrt{Q/n}}$

$$T \sim t_n$$

  - $E(T) = 0$ and Var$(T) = n/(n-2)$ for $n > 2$
  - Shape of the t-distribution is similar to the normal distribution but more spread out (heavier tails)
- As sample size increases, the t-distribution approaches the standard normal distribution

Sampling Distributions

The distribution of the sample statistic (e.g., sample mean, sample variance, regression coefficients) over repeated independent sampling

- **Sampling variability**: The variability of the sample statistic across different samples
- **Standard error**: The estimate of the standard deviation of the sampling distribution of a sample statistic
- **Central Limit Theorem (CLT)**: Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample of size n from a population with mean μ and variance σ^2. Then, as n approaches infinity, $\bar{X} \sim N(\mu, \sigma^2/n)$ so that $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- **Recall**:
  - The sampling distribution of the *sample proportion* $\hat{p}$ approaches a normal distribution with mean p and variance $p(1-p)/n$
  - The sampling distribution of *sample means* follows a t-distribution with $n-1$ degrees of freedom

Hypothesis Testing

- **Null hypothesis** H_0: A statement about the population parameter that is assumed to be true and test against an alternative hypothesis H_1
- **Type I error**: Rejecting the H_0 when it is true
- **Type II error**: Failing to reject the H_0 when it is false
- **Significance level**: The probability of committing a Type I error, denoted by α
- **Confidence interval**: A range of values where we are $1 - \alpha$ confident that the true population parameter lies within
  - estimate $\pm$ critical value $\times$ standard error
  - What happens if the CI contains the hypothesized value?
- Tests can only allow us to reject the H_0 or fail to reject the H_0 but never accept the H_0. Why?

Hypothesis Testing (cont.)

- **p-value**: The probability of observing the sample (statistic) if H_0 were true
  - What does it mean when p-value is less than α?
- **Critical value**: The value that separates the rejection region from the non-rejection region in hypothesis testing
  - What is critical value for a two-tailed test with $\alpha = 0.05$?
- **Critical region**: The range of values that leads to the rejection of the H_0
- **One-tailed test**: A hypothesis test that tests the H_0 in one direction
- **Two-tailed test**: A hypothesis test that tests the H_0 in both directions