## Probability.

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

$P(A,B) = P(A|B) \cdot P(B)$

If A and B are independent:

$P(A|B) = P(A)$. $P(B|A) = P(B)$

Bayes's Rule: $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

### Expected Value:

For a discrete random variable:

$$E[X] \equiv \sum_{-\infty}^{+\infty} x \cdot P[X = x]$$

Continuous random variable:

$$E[X] \equiv \int_{-\infty}^{+\infty} x \cdot p(x)\, dx$$

"E" tells us what to expect from the average of many observations of the random variable.

**Variance**: $Var(X) = E[(X - \bar{X})^2]$ $\sigma^2$

**Standard deviation**: $\sigma$.

### Distributions.

· Bernoulli Trials: Binary outcome. Discrete

· Independent · Fixed probability: $p$ and $(1-p)$.

$P(x) = p^x (1-p)^{(1-x)}$

$P(x) = \begin{cases} 1-p & \text{if } x=0 \\ p & \text{if } x=1 \end{cases}$ Mean: $p$ Variance: $1-p$

The Poisson process: Continuous process.

· Constant average rate. denoted as $\lambda$

Geometric Distribution: what is the probability it takes k trials to obtain the first success?

PDF: $P[X=k] = p(1-p)^{k-1}$. $k \geq 1$.

Mean $= \frac{1}{p}$. $\sigma^2 = \frac{1-p}{p^2}$

Binomial Distribution

$P[X=k] = \binom{N}{k} p^k (1-p)^{N-k}$

Exponential Distribution

Success occurs in a fixed rate measuring the until a success occurs.

PDF: $p(x) = \lambda e^{-\lambda x}$.

Mean: $\frac{1}{\lambda}$. variance: $\frac{1}{\lambda^2}$

$F(x) = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}$

Poisson Distribution: How many success occur in a fixed amount of time.

$P[k \text{ successes in time } T] = (\lambda T)^k \frac{e^{-\lambda T}}{k!}$

## Uniform Distribution

When the time an event occurs is uniformly distributed. the number of events in a time interval is Poisson Process.

## The Gaussian Distribution

Standard PDF: $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$\mu = 0$ and $\sigma = 1$

For mean $\mu$ and $\sigma^2$:

$p_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

If $X_1$ and $X_2$ are Gaussian, then $X_3 = aX_1 + bX_2$ is Gaussian.

## Joint Distributions.

Marginal. $P(x) = \sum_y P(X, Y=y)$

$P(x) = \sum_Y P(X, Y)$

Independence: If $P(A|B) = P(A)$ then: $P(A,B) = P(A) \cdot P(B)$

## Correlation

Covariance: Describe a relationship between random variables with a single number

$Cov(X_1, X_2) = E[(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)]$

Correlation coefficient:

$\rho(X_1, X_2) = \frac{E[(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)]}{\sigma_{X_1} \sigma_{X_2}}$

Value between -1 and 1.

$\rho = 0 \rightarrow X_1$ and $X_2$ is uncorrelated.

$>0 \rightarrow$ positively correlated.

$<0 \rightarrow$ anticorrelated.

## Functions of Random Variables.

Linearity of Expectation.

$E[X+Y] = E[X] + E[Y]$.

Suppose we have n random variables with the same distribution.

$E[\sum X_i] = n E[X]$.

Each has mean $\mu$, then the mean of the sum is $n\mu$.

$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$

$= E[XY] - E[X] \cdot E[Y]$.

$Var(X) = E[X^2] - E[X]^2$.

## Variance of a Sum

$Var(X+Y)$

$= E[(X+Y)(X+Y)] - E[(X+Y)]^2$

$= Var(X) + Var(Y) + 2Cov(X,Y)$.

When adding independent random variables, variance sum.

Summing n independent random variables each with mean $\mu$ and variance $\sigma^2$.

The sum has mean $n\mu$ and variance $n\sigma^2$.

## The Central Limit Theorem.

Bernoulli trials: Mean: $p$ with probability $p$ Std: $\sqrt{p(1-p)}$

In CLT: $\mu = p$ and $\sigma = \sqrt{\frac{p(1-p)}{n}}$

Sum of the Bernoulli Trials.

$\mu = np$ and $\sigma = \sqrt{np(1-p)}$

68. 95. 99.7 Rule

68% of all values are within 1 SD of the mean

95% are within 2 SD.

99.7% are within 3 SD.

Standard error: $\frac{\sigma^2}{n}$

Confidence. 95% Confidence CI.

$P(\hat{\mu} - f \leq \mu \leq \hat{\mu} + f) \geq 0.95$

Gaussian Case: Distribution of sample means:

$\hat{\mu} \sim N(\mu, \frac{\sigma^2}{n})$.

$(\hat{\mu} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \hat{\mu} + 1.96 \cdot \frac{\sigma}{\sqrt{n}})$.

Sample Mean: $\hat{\mu} = \frac{1}{n} \sum X_i$

Unbiased sample standard deviation:

$S = \sqrt{\frac{1}{n-1} \sum_i (X_i - \hat{\mu})^2}$

Test statistic times standard error.

$f_{95} = \begin{cases} 1.96 \cdot \frac{S}{\sqrt{n}} & \text{if } n > 30 \\ t_{0.975}(n-1) \cdot \frac{S}{\sqrt{n}} & \text{if } n \leq 30 \end{cases}$

$CI = (\hat{\mu} - f_{95}, \hat{\mu} + f_{95})$

Conservative estimate:

$\frac{1}{2}$ for the S

2 for the 1.96

The 95% confidence interval for the Bernoulli is approximately $\hat{\mu} \pm \frac{1}{\sqrt{n}}$

## Hypothesis Testing.

· Type I Error: False positive.

"mistaken rejection of the $H_0$".

Control the false positive rate to avoid making incorrect "discoveries".

· Type II Error: False negative.

"Concluding there is no difference when in fact there is a difference."

Reject the $H_0$ if $p < \alpha$.

Z-score: standard score. $Z = \frac{X - \mu}{\sigma}$

t-statistic: $t = \frac{\hat{\mu} - \mu}{\frac{S}{\sqrt{n}}}$

The family-wise error rate (FWER).

Making no false positive errors at all:

$P(V>0) = 1 - P(\text{no rejection of any of } m \text{ nulls})$.

For a fixed p-value threshold $\alpha$:

$P(V>0) = 1 - P(\text{no rejection of any of } m \text{ nulls})$

$= 1 - (1-\alpha)^m$.

Choose the per hypothesis threshold

$\alpha = \frac{\alpha_{FWER}}{m}$

Divide the p-value threshold you would normally use by the total number of hypotheses being tested to control the FWER.

Bonferroni correction is a good conservative method to control FWER when testing many hypotheses.

## Types of Distribution.

· Geometric: Tracks how many attempts (failures) it takes to achieve a desired outcome (success).

· The number of times you need to flip a coin until you get heads.

Binomial distribution: Predicts the likelihood of a specific outcome (success or failure) in situations with only two possible results. | The probability of making a free throw in basketball.

Exponential Distribution: Estimates the time

· Estimates the time you have to wait until something happens.

· The time between arrivals of buses at a bus stop.

Poisson Distribution

· The # of rare events occurring in a

fixed interval of time or space

· The # of customers arrivals at a store in an hour.

**Uniform Distribution**: Ensure that every where in a range is equally likely. Ex.: Rolling a fair six-sided die, where each number has a $\frac{1}{6}$ chance of occurring.

**Gaussian Distribution**: Describes data where most values cluster around the mean. | The heights of people in a population, forming a bell-curve.

**Notes on Integral**

$$\int_a^b f(x)\,dx = F(b) - F(a) \qquad \int x^n dx = \frac{x^{n+1}}{n+1} + C \quad n \neq -1$$

Sample question.

Find $E[X^2]$ given the PDF: $f(x) = \begin{cases} x+1, & 3 < x < 7 \\ 0, & \text{otherwise.} \end{cases}$

$$\int_3^7 x^2(x+1)\,dx = \int x^3 + x^2\,dx = \frac{2056}{3}$$

**Simple Random Sampling** $\{20, 22, 24\}$ Where all possible samples of size two.

Mean: $\frac{20+22+24}{3} = 22$ $\quad \sigma^2 = \frac{400+484+576}{3} - 484 = 2.6$

Formula of variance for population: $\sigma^2 = \frac{\Sigma(x_i-\mu)^2}{N}$

**Mean of a random sample.**

$\mu = 80$, $\sigma = 6$, $n = 36$, sample mean $(\bar{x}) = 48$

Standard deviation of the sample mean (Standard Error)

$SE = \frac{6}{\sqrt{n}}$.

**Confidence Interval for the population mean** $\mu$ with a known population standard deviation $\sigma$:

$CI = \bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$

95% CI is critical value $= 1.96$

Z-score for the 95th percentile is 1.645.

Top 5% $\quad 1.645 = \frac{X-75}{10}$, $X = 91.45$

**CI for Sample Mean**

$CI = SM \pm (\text{Critical value} \times \frac{\text{Sample SD}}{\sqrt{\text{Sample size}}})$

i.i.d: Independent: Data points are independent if the occurrence or value of one data point does not affect the occurrence or value of another. They are unrelated.

· Identically distributed: Data points are identically distributed if they all follow the same probability distribution with the same parameters.

Ex. Flipping a coin multiple times.

Point estimation varies from sample to sample.

Bias: Measures the difference between the expected value of the estimator and the true value of the parameter.

Variance: Measures how much the estimator can vary as a function of the data sample.

Trade-off: Method with lower bias tend to have higher variance and vice versa.

**Steps of Finding MLE.**

1. Find the likelihood function (joint probability).
2. Find the log-likelihood function
3. Find the derivative of log-likelihood function.
4. Set it equal to 0 and solve.

**Probability:**

$P(\text{at least one success}) = 1 - P(\text{no success})$

Binomial distribution: $P_k = \binom{n}{k} p^k (1-p)^{n-k}$

Poisson distribution: $P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$

**Variance Calculation via E[X]**

$$Var(X) = E[X^2] - (E[X])^2$$

**CLT and Variance.**

Find the variance for the average of the 100 samples.

$Var(\bar{X}) = \frac{Var(X)}{n}$

**Covariance**

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

**Correlation coefficient.**

$$r = \frac{Cov(X,Y)}{S_x \cdot S_y}$$

$$S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$S_x = \sqrt{\frac{(1000-1150)^2 + (1050-1150)^2 + (1150-1150)^2 + (1400-1150)^2}{4}}$$

**E[X] Examples Cal.**

Continuous random variable $f(x) \begin{cases} x, & \text{if } 0 \le x \le 1 \\ x-1, & \text{if } 1 \le x \le 2 \\ 0, & \text{if otherwise.} \end{cases}$

Find $E[X^3]$.

$\int_{-\infty}^{\infty} x^3 f(x)\,dx$ Into two parts.

For $0 \le x \le 1$: $f(x) = x$: $\int_0^1 x^3 \cdot x\,dx$

For $1 \le x \le 2$: $f(x) = x-1$: $\int_1^2 x^3 \cdot (x-1)\,dx$.

Ex. $(X,Y)$ be a random vector, joint PDF:

$f_{X,Y}(x,y) \begin{cases} 4xy, & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0, & \text{if otherwise.} \end{cases}$ Find $E[X(Y^2)]$

$E[X(Y^2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(xy^2) \cdot 4xy\,dx\,dy$

$E[X(Y^2)] = \int_0^1\int_0^1 (xy^2) \cdot 4xy\,dx\,dy = \frac{1}{3}$

**Inverse Z-score function.**

$X = \mu + Z \cdot \sigma$

**t-distribution CI**

$\bar{x} \pm t_{\frac{\alpha}{2}, df} \cdot (\frac{s}{\sqrt{n}})$

As the sample size increases, the width of the CI decreases.

**One-tailed hypothesis test (t)**

$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

t statistic > critical value

Reject $H_0$.

**Type I and Type II Errors.**

Type I: Rejecting $H_0$ when it's actually true (False positive).

Type II: Failing to reject $H_0$ when $H_1$ is actually true.

**Log Differentiation Table.**

$\log ab = \log a + \log b$

$\log \frac{a}{b} = \log a - \log b$

$\log a^n = n \log a$

$\log e^x = e^{\log x} = x$

$\frac{d}{dx}\log x = \frac{1}{x}$

$\frac{d}{dx}(uv) = uv' + vu'$

$\frac{d}{dx}(\frac{u}{v}) = \frac{vu' - uv'}{v^2}$

$\frac{d}{dx}z(y(x)) = \frac{dz}{dy}\frac{dy}{dx}$.

**Determine $\frac{dy}{dx}$ for function**

$y = f(x)$

$\frac{dy}{dx} = uv' + u'v$

$\frac{d}{dx}\ln(y) = \frac{1}{y} \cdot \frac{dy}{dx}$

**Example for Bias Variance and Mean.**

$\{y^{(1)}, y^{(2)}, \dots y^{(n)}\}$ i.i.d sample continuous uniform distribution. Parameters 0 and $\theta$.

$\hat{\theta} = \frac{2}{n}(y^{(1)}, y^{(2)} \dots y^{(n)})$ be an estimator of $\theta$.

Mean: $\frac{a+b}{2}$ $\quad$ Variance: $\frac{(b-a)^2}{12}$

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

$E[\hat{\theta}] = E[\frac{1}{n}\sum_{i=1}^{n}x^{(i)}]$

$= \frac{1}{n}\sum_{i=1}^{n} \cdot E[x^{(i)}]$

$= \frac{1}{n}\sum_{i=1}^{n} \cdot \frac{1}{\lambda}$

$= \frac{1}{n} \times n \cdot \frac{1}{\lambda}$

$= \frac{1}{\lambda}$ $\quad$ Bias: $E[\hat{\theta}] - \theta$

$= \frac{1}{\lambda} - \frac{1}{\lambda} = 0$

Unbiased.

Mean of $\hat{\theta}$ 1. Find its expected value $E(\hat{\theta})$

The mean $E(y^{(i)})$ is: $\frac{0+\theta}{2} = \frac{\theta}{2}$ Expected value:

$E(\sum_{i=1}^{n} y^{(i)}) = nE(y^{(i)}) = n \cdot \frac{\theta}{2} = \frac{n\theta}{2}$

The expected value $E(\hat{\theta})$ is $E(\hat{\theta}) = \frac{2}{n} \cdot$

$E(\sum_{i=1}^{n} y^{(i)}) = \frac{2}{n} \cdot \frac{n\theta}{2} = \theta$, the mean of $\hat{\theta} = \theta$.

Variance: $Var(y^{(i)}) = \frac{(\theta-0)^2}{12} = \frac{\theta^2}{12}$

Since the observations are independent, the variance of the sum is the sum of variance.

$Var(\sum_{i=1}^{n} y^{(i)}) = n \cdot \frac{\theta^2}{12} = \frac{n\theta^2}{12}$. $\quad Var(cX) = c^2 Var(X)$.

$Var(\hat{\theta}) = (\frac{2}{n})^2 Var(\sum_{i=1}^{n} y^{(i)}) = \frac{4}{n^2} \cdot \frac{n\theta^2}{12} = \frac{\theta^2}{3n}$

**Maximum Likelihood**

Example: Will spins a coin 3 times and observes no heads. He then gives the coin to Amelia. She spins it until the first head occurs, and ends up spinning it four times total. Let $\theta$ denote the probability the coin comes up heads. Likelihood function and MLE.

$X = \{T, T, T, T, T, T, H\}$

Compute the likelihood function of $\theta$.

Step 1: $p(X;\theta) = (1-\theta)\cdot(1-\theta)\dots(1-\theta)\theta = (1-\theta)^6\theta$

Step 2: $\log p(X;\theta) = \log((1-\theta)^6\theta)$

$= \log((1-\theta)^6) + \log(\theta) = 6\log(1-\theta) + \log\theta$

Step 3: $\frac{\partial}{\partial\theta}\log p(X;\theta) = \frac{\partial}{\partial\theta}[6\log(1-\theta) + \log\theta]$

$= -\frac{6}{1-\theta} + \frac{1}{\theta}$

Step 4: $-\frac{6}{1-\theta} + \frac{1}{\theta} = 0$, $\hat{\theta} = \frac{1}{7}$

**Bias Revisited**: To determine if $\hat{\theta}$ is a biased estimator, we need to compare its expected value $E(\hat{\theta})$ with the true population mean $\frac{1}{\lambda}$.

The bias of an estimator $\hat{\theta}$ is: $Bias(\hat{\theta}) = E[\hat{\theta}] - \theta$

Now: $E[\hat{\theta}] = E[\frac{1}{n}\sum_{i=1}^{n} E[x^{(i)}]]$ Because expectation is a linear operator: $E[\hat{\theta}] = \frac{1}{n}\sum_{i=1}^{n} E[x^{(i)}] = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda}$

$= \frac{1}{n} \times n \times \frac{1}{\lambda} = \frac{1}{\lambda}$ $\quad Bias = E[\hat{\theta}] - \theta = \frac{1}{\lambda} - \frac{1}{\lambda} = 0$

**Linearity of Expectation.** $E[aX + bY] = aE[X] + bE[Y]$ The expectation of a sum of random variables is equal to the sum of their expectations.

**Law of the total probability.**

$P(B) = \sum_{i=1}^{n} P(B|A_i) \times P(A_i)$

$P(B|A_i)$ is the probability of B given A.

**Bayes's Rule.**

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

**Sum of Variance.**

$n \cdot E[X] \pm \sqrt{n \cdot Var(X)}$.

For range problem.

**Sample Variance**

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

**Expected value of the Sample Mean:** $E(\bar{X}) = \mu$

**How many meals:** $n \cdot E[X] = \cancel{\text{Tw}}$