Prob #1

Set Notation

- A set is a collection of objects or "elements"

$A = \{a_1, a_2, a_3\}$

- The size of a set is the number of elements in that set
- There are special sets
  + S: The sample space or everything that is possible
  + $\varnothing$: The null space
- Operations
  + $\cup$ - Union: Elements in either set (or)
  + $\cap$ - Intersection: Elements in both sets (and)
  + $\subset$ - Inside of: Obvious
  + $\in$ - Element of
  + $\times$ - Cartesian Product: New set with element pairs
- A set does not worry about order
  + But a sequence does

Combinatorial Methods

- These are methods for finding the size of a set given details about it
- Each one is based on different assumptions about the situation

1) Factorial $(n!)$
   + represents the number of ways a set of size n can be arranged
   + $n \cdot (n-1) \ldots$

2) Permutation $({}_nP_r)$
   + Number of ways to put n objects in r positions
   + ${}_nP_r = \frac{n!}{(n-r)!}$
   + Cares about order

3) Combination $\left({}_nC_r \,/\, \binom{n}{r}\right)$
   + Number of ways to pick r objects from n objects
   + ${}_nC_r = \frac{n!}{r!(n-r)!}$
   + Doesn't care about order
   + Represents the coefficients of a binomial

Probabalistic Laws

- There are a few laws which describe how how the operations on a set affects its probability
- Conditional

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

Defines Independance if $\mathbb{P}(B|A) = \mathbb{P}(B)$

- Additive Law

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

- Multiplicative

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

- Bayes Rule
  + $\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}{\mathbb{P}(A)}$
  + B forms a partition with i parts
  + j refers to a specific partition
  + Allows you to update a probability given new data

Sample point Method

1) Describe the space
2) Figure out the number in that space
3) Describe the simple events
4) Use simple events to describe event

4) Partition $\left({}_nC_{r_1, r_2 \ldots} \,/\, \binom{n}{r_1\, r_2\, r_3 \ldots}\right)$
   + Number of ways you can divide n objects into groups of size $r_1, r_2, r_3 \ldots$
   + ${}_nC_{r_1, r_2} = \frac{n!}{r_1!\, r_2!\, r_3! \ldots}$
   + Doesn't care about the order of the partitions
   + The sum of the partitions must equal the size of the set
   + Represent multinomial coefficients

Approach Suggestions

- Use symmetry

Problem Approach

- There are two major structured methods for probability

Event composition

1) Describe the sample space
2) Describe the wanted probability in terms of known probabilities
3) Apply probability laws

Prob #2

- Random variables
  - + Expectation
  - + Variance
  - + Other operations
- Distributions
  - + Binomial
  - + Geometric
  - + Hypergeometric
  - + Poisson

Random Variables

- Variables that take on some specific values according to some distribution
- Usually described by a piece wise or a set
- The important aspect is the distribution they have which describes various aspects about them
- Some important properties
  - + Expectation: $\mathbb{E}Y = \sum_{y \in S} y \, \mathbb{P}(Y=y)$
    - Describes the mean
    - Is linear in nature
  - + Variance: $V(Y) = \mathbb{E}(Y - \mathbb{E}Y)^2 = \mathbb{E}Y^2 - (\mathbb{E}Y)^2$
    - Describes how spread out the distribution is
    - Additive if the r.v.s are independent and adding
  - + Standard deviation: $\sigma(Y) = \sqrt{V(Y)}$
- Any function of Y just transforms the values of Y

General distributions

- Binomial: $Y \sim Bin(n, p)$
  - + Only two possible states
  - + One state in n slots with probability p
  - + Events are independent
  - + $\mathbb{P}(Y=y) = \binom{n}{y} p^y (1-p)^{n-y}$
  - + $\mathbb{E}(Y) = np$
  - + $V(Y) = np(1-p)$
- Geometric: $Y \sim Geo(p)$
  - + Probability of lasting until
  - + Series of failures until success
  - + $\mathbb{P}(Y=y) = (1-p)^{y-1} p$
  - + $\mathbb{E}(Y) = \frac{1}{p}$ + $V(Y) = \frac{1-p}{p^2}$
- Hyper Geometric: $Y \sim HyperGeo(N, r, n)$
  - + Similar to binomial only if the probability is affected by drawing marked objects
  - + Works if the number of marked values is of similar size to total number
  - + $\mathbb{P}(Y=y) = \dfrac{C^r_y \, C^{N-r}_{n-y}}{C^N_n}$
    - + N: Total num
    - + r: mark num
    - + n: draw num
  - + $\mathbb{E}(Y=y) = \frac{nr}{N}$ + $V(Y) = \frac{nr}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}$
- Poisson: $Y \sim Pois(\lambda)$
  - + Any number of events can happen during the time period
  - + λ: rate over time period
  - + $\mathbb{P}(Y=y) = e^{-\lambda} \cdot \frac{\lambda^y}{y!}$
  - + $\mathbb{E}(Y) = \lambda$
  - + $V(Y) = \lambda$

Prob #3

- Continuous R.V.
  + Properties
  + CDF
  + PDF
  + Expectation
  + Variance
- Types of Continuous R.V.
  + Indicator
  + Uniform
  + Normal
  + Exponential
  + Gamma
  + Beta

Uniform

- Time for the arrival of a poisson object
- pdf: $\frac{1}{\theta_2 - \theta_1} \mathbb{1}_{[\theta_1, \theta_2]}(y)$
- cdf: $\frac{y - \theta_1}{\theta_2 - \theta_1} \mathbb{1}_{[\theta_1, \theta_2]}(y) + \mathbb{1}_{[\theta_2, \infty)}(y)$
- $\mathbb{E}\,\mathcal{U}(\theta_1, \theta_2) = \frac{\theta_1 + \theta_2}{2}$

  $V(\mathcal{U}(\theta_1, \theta_2)) = \frac{(\theta_2 - \theta_1)^2}{12}$

Normal

- pdf: $f(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)$
- $\mathbb{E}(Y) = \mu$
- $V(y) = \sigma^2$
- Affline transformations: $Z := a + bY \sim N(a + b\mu, b^2\sigma^2)$

Exponential

- Time until first poisson event
- pdf: $f(y) = \lambda e^{-\lambda t}$
- cdf: $F(y) = 1 - e^{-\lambda t}$ $\quad t \in [0, \infty)$
- $\mathbb{E}(Y) = \frac{1}{\lambda}$ $\quad V(Y) = \frac{1}{\lambda^2}$
- Scalar invariance: $\tau \sim Exp(a)$ then $\tau/\mu \sim Exp(\lambda\mu)$
- Memoryless: $\mathbb{P}(t > s + \tau \mid t > \tau) = \mathbb{P}(t > s)$
- $\min\{\tau_1, \tau_2 \cdots\} \sim Exp(\lambda)$

Gamma

- Represents sum of exponentials

  $\tau_1 + \tau_2 \cdots + \tau_n \sim Gamma(n, 1/\lambda)$
- $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1}\, dy$
- pdf: $f(y) = e^{-t/\beta} \frac{t^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)}$
- $\mathbb{E}Y = \alpha\beta$ $\quad V(y) = \alpha\beta^2$
- α controls left while β controls right
- χ^2 is a special case of Gamma
  + $\chi^2(v) \sim Gamma(v/2, 2)$
  + v is the degree's of freedom

Beta

- $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$
- pdf: $f(y) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)} \mathbb{1}_{[0,1]}(y)$
- If α and β are integers then

  $F(y) = \sum_{i=\alpha}^{n} C_i^n y^i (1-y)^{n-i}$

  $n := \alpha + \beta - 1$
- $\mathbb{E}Y = \frac{\alpha}{\alpha + \beta}$ $\quad V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha + \beta + 1)}$

Conditional Distributions

$$\mathbb{P}(X \in dx \mid Y = y) := \frac{\mathbb{P}(X \in dx, Y \in dy)}{\mathbb{P}(Y \in dy)}$$

$$\mathbb{E}(X \mid Y = y) = \int_{\mathbb{R}} x \, \mathbb{P}(X \in dx \mid Y = y)$$

Properties

1) If X is constant $\Rightarrow \mathbb{E}(X \mid Y) = c$
2) If Y is a constant with probability $1 \Rightarrow \mathbb{E}(X \mid Y) = \mathbb{E}X$
3) Linearity $\Rightarrow \mathbb{E}(\lambda X + \mu Z \mid Y) = \lambda \mathbb{E}(X \mid Y) + \mu \mathbb{E}(Z \mid Y)$
4) Monotonicity $\Rightarrow$ if $X \le Z$ with probability one
   $\mathbb{E}(X \mid Y) \le \mathbb{E}(Z \mid Y)$
5) Tower property $\Rightarrow \mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}X$
6) If Z is a deterministic function of Y then
   $\mathbb{E}(XZ \mid Y) = Z \cdot \mathbb{E}(X \mid Y)$
7) For any convex function ϕ we have
   $\mathbb{E}(\phi(X) \mid Y) \ge \phi(\mathbb{E}(X \mid Y))$
   Also works for regular expectations
   $\mathbb{E}\phi(X) \ge \phi(\mathbb{E}\phi)$

General Definitions

- $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\, dx\, dy = \mathbb{P}(X < x, Y < y)$ — cumulative
- conditional — $\mathbb{P}(X \mid Y) = \dfrac{\mathbb{P}(X, Y)}{\mathbb{P}(Y)}$
- marginal — $\mathbb{P}(Y) = \int_{-\infty}^{\infty} \mathbb{P}(X, Y)\, dx$

Independance

X and Y are indep iff   if indep $V(X+Y) = V(X) + V(Y)$

1) $\mathbb{P}(X, Y) = \mathbb{P}(X)\mathbb{P}(Y)$
2) $F(X, Y) = F(X)F(Y)$
3) $f(X, Y) = f(X) f(Y)$
4) $\mathbb{P}(X \mid Y) = \mathbb{P}(X)$
5) $\mathbb{E}[f(X) g(Y)] = \mathbb{E}f(X)\, \mathbb{E}g(Y)$

Indep measures

- Covariance: $\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = Cov(X,Y)$
  + $Cov(aX, bY) = ab\, Cov(X,Y)$
- correlation: $\rho(X,Y) = \dfrac{cov(X,Y)}{\sqrt{V(X)V(Y)}}$

Reference Table

Name	Symbol	pdf/pmf	cdf/cmf	mgf	mean	Variance
Bernouli	$B(p)$	$p^k(1-p)^k$	$\mathbb{1}_{[p,1]}(1-p)(k) + \mathbb{1}_{[1,p]}(k)$	$q + pe^t$	p	$p(1-p)$
Binomial	$Bin(n,p)$	$\binom{n}{k} p^k q^{n-k}$	Sum pmf	$(q + pe^t)^n$	np	$np(1-p)$
Geometric	$Geom(p)$	$(1-p)^{k-1} p$	$1-(1-p)^k$	$\dfrac{pe^t}{1-(1-p)e^t}$ denom>0	$1/p$	$1-p/p^2$
Poisson	$Pois(\lambda)$	$\dfrac{\lambda^k e^{-\lambda}}{k!}$	$e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \dfrac{\lambda^i}{i!}$	$e^{\lambda(e^t - 1)}$	λ	λ
Exponential	$Exp(\lambda)$	$\lambda e^{-\lambda t}$	$1 - e^{-\lambda t}$	$1/(1 - t/\lambda)$ $t<\lambda$	$1/\lambda$	$1/\lambda^2$
Uniform	$U(\theta_1, \theta_2)$	$1/\theta_2 - \theta_1$	$\dfrac{y - \theta_1}{\theta_2 - \theta_1}$	$\dfrac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$	$\dfrac{\theta_2 + \theta_1}{2}$	$\dfrac{(\theta_2 - \theta_1)^2}{12}$
Normal	$N(\mu, \sigma^2)$	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$	$e^{\mu t + \sigma^2 t^2/2}$	μ	σ^2
Gamma	$\Gamma(a, \beta)$	$e^{-t/\beta} \dfrac{t^{a-1}}{\beta^a \Gamma(a)}$		$(1 - t\beta)^{-a}$ $t<1/\beta$	$a\beta$	$a\beta^2$
Beta	$Beta(a, \beta)$	$\dfrac{y^{a-1}(1-y)^{\beta-1}}{\beta(a,\beta)} \mathbb{1}_{[0,1]}(y)$			$\dfrac{a}{a+\beta}$	$\dfrac{a\beta}{(a+\beta)^2(a+\beta+1)}$
χ^2	$\chi^2(v) \sim \Gamma(v/2, 2)$	See gamma			v	$2v$

Stats #1

Estimators

- Estimators are ways of approximating the parameters a distribution
  - They are a r.v. in and of themselves
- Each estimator has various properties
  - Bias $:= B(\hat{\theta}) = \hat{\theta} - \theta$
  - Error $:= |\hat{\theta} - \theta|$
  - Mean square error $:= MSE(\hat{\theta}) = \mathbb{E}|\hat{\theta}-\theta|^2 = V(\hat{\theta}) + B(\hat{\theta})^2$
- Some properties depend on having a large sample size due to the law of large numbers and the central limit theorem. These are asymptotic properties
  - Asymtotic normality: $\sqrt{n}(\hat{\theta}_n - \theta) \to N(0, \sigma^2)$
    - This implies consistency

Popular estimators

- Sample mean $:= \bar{Y} = \frac{1}{n}\sum_{i=0}^{n} Y_i$
  - Unbiased for the mean
  - Asymtotically normal
  - Variance: $\frac{\sigma^2}{n}$
- Sample variance $:= S^2 = \frac{1}{n-1}\sum_{i=0}^{n}(Y_i - \mu)^2$
  - Unbiased estimator for σ^2
  - Follows a χ^2 distribution

Parameter	Point Estimator	$\mathbb{E}(\hat{\theta})$	Standard error
μ	$\bar{Y}$	μ	$\sigma/\sqrt{n}$
p	$\hat{p} = Y/n$	p	$\sqrt{pq/n}$
$\mu_1 - \mu_2$	$\bar{Y}_1 - \bar{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

Important Laws

- Law of large numbers: The expected value from an experiment gets closer to the real expected value
- Central limit theorem: As $n \to \infty$ any distribution approaches a normal distribution

Confidence Intervals

- Pivot method
  - Create a function of Y that has a known distribution then use that to describe Y
- Common pivots
  - $\frac{\sqrt{n}(Y - \bar{Y})}{\sigma} \sim N(0,1)$: Known σ
    - $\mu \in \left(\bar{Y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{Y} + z_{\alpha/2}\sigma/\sqrt{n}\right)$
  - $\frac{\sqrt{n}(Y - \bar{Y})}{S} \sim T(n-1)$: Both unknown
    - $\mu \in \left(\bar{Y} - S t_{\alpha/2}/\sqrt{n}, \bar{Y} + S t_{\alpha/2}/\sqrt{n}\right)$
  - $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
    - $\sigma \in \left(\frac{n-1}{\chi^2_{\alpha/2}} s^2, \frac{n-1}{\chi^2_{1-\alpha/2}} s^2\right)$
  - $\bar{Y} - \bar{Z} - (\mu_1 - \mu_2) \sim N(0, \sigma_1^2/n_1 + \sigma_2^2/n_2)$
    - $\mu_1 - \mu_2 \in \bar{Y} - \bar{Z} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \sigma_2^2/n_2}$
    - Assumes σ_1 and σ_2 are known

- Making estimators
  - Comparisons
    - Efficiency
    - Sufficiency
    - Completeness
  - Best estimators
    - MVUE
    - MLEs
    - Moment matching estimators

Making Estimators

- Relative efficiency $:= \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$
- Sufficiency: $\mathbb{P}(\{Y_i = y_i\} | \hat{\theta})$ does not depend on θ
- Rao-blackwell: $\hat{\theta}^* := \mathbb{E}(\hat{\theta} | U)$
- Likelihood func: $L = \prod_1^n f(y_i; \theta)$
- Fisher-Neyman: Decomposes the likelihog
- If g is monotone and U suff. then $g(U)$ is also sufficient
- If the likelihood ratio does not depend on θ then $\hat{\theta}$ is sufficient and complete
- Finding MLE
  - Take log of L
  - differentiate and solve for zero
- Can use this to find a general confidence interval finder

$$z := \frac{\sqrt{-\mathbb{E}^{\theta}\partial^2_{\theta\theta}\log f(Y_i;\theta)}}{|G'(\theta)|}\sqrt{n}\left(G(\hat{\theta}) - G(\theta)\right) \sim N(0,1)$$

- Moment matching estimator

$$m_k(\hat{\theta}) = \mathbb{E}^{\theta} Y^k = \frac{1}{n}\sum_{i=1}^{n} Y^k$$

Lecture #4

- Relative efficiency
- Sufficiency
- Completeness
- Minimum variance unbiased estimator (MVUE)

Lecture #5

- Moment matching estimators
- Maximum likelihood estimator (MLE)