1 STAT 400, Aryn Harmon

1.1 Midterm 1 Material

Probability is a real-valued function: 1. $\mathbb{P}(S) = 1$; 2. $\mathbb{P}(A) \geq 0$; 3. If A_1, A_2 are mutually exclusive events, $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$ and so on.
Inclusion-Exclusion: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
Conditional Probability: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ given $\mathbb{P}(B) > 0$. $\mathbb{P}(A|B) \neq \mathbb{P}(A)$ unless A and B are independent. Also see Bayes's Theorem. Probability of a string of unions given B is equal to the sum of the individual conditional probabilities.
Multiplication Rule: Probability of two events both occurring: $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A)$ or $\mathbb{P}(A \cap B) = \mathbb{P}(B) \cdot \mathbb{P}(A|B)$ (one is easier than the other).
Bayes's Rule: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A) \cdot \mathbb{P}(B|A)}{\mathbb{P}(B)}$. $\mathbb{P}(A)$ is the *prior probability* of A. $\mathbb{P}(A|B)$ is the *posterior probability* of A given that B occurred. Use to invert probabilities.
Bayes's Rule 2: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A) \cdot \mathbb{P}(B|A)}{\mathbb{P}(A) \cdot \mathbb{P}(B|A) + \mathbb{P}(A') \cdot \mathbb{P}(B|A')}$.
Bayes's Rule Full: Given some partition of S: $A_1 \cup \cdots \cup A_k = S$. $\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i) \cdot \mathbb{P}(B|A_i)}{\sum_{m=1}^{k} \mathbb{P}(A_m) \cdot \mathbb{P}(B|A_m)}$.
Independence: $\mathbb{P}(A|B) = \mathbb{P}(A)$, $\mathbb{P}(B|A) = \mathbb{P}(B)$, and $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.
Pairwise Independence: All of the following must hold: $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C)$, and $\mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C)$.
Mutual Independence: A, B, and C must be pairwise independent and in addition: $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$.
Number of ways to take r objects from n candidates:

	Ordered Sample	Unordered Sample
W. R.	n^r	$\binom{n+r-1}{r}$
W/o. R.	${}^nP_r$	$\binom{n}{r}$

${}^nP_k = \frac{n!}{(n-k)!}$ - permutation
$\binom{n}{k} = {}^nC_k = \frac{n!}{k!(n-k)!}$ - combination

Binomial Theorem: $(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}$. $\binom{n}{n-r} = \binom{n}{r}$, $\binom{n}{0} = \binom{n}{n} = 1$, $\binom{n}{1} = \binom{n}{n-1} = n$.
Some magic formulae: 1. $\sum_{r=0}^{n} \binom{n}{r} = 2^n$. 2. $\sum_{r=0}^{n} (-1)^r \binom{n}{r} = 0$. 3. $\sum_{r=0}^{n} \binom{n}{r} p^r (1-p)^{n-r} = 1$.
Pascal's Equation: $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$.
Hypergeometric Distribution: $f(x) = \mathbb{P}(X = x) = \frac{\binom{N_1}{x} \cdot \binom{N_2}{n-x}}{\binom{N}{n}}$, where $x \leq n$, $x \leq N_1$, $n - x \leq N_2$ and $N = N_1 + N_2$. $\mathbb{E}(X) = n \frac{N_1}{N}$ and $Var(X) = n \cdot \frac{N_1}{N} \cdot \frac{N_2}{N} \cdot \frac{N-n}{N-1}$. Example: Urn model: N_1 red balls, N_2 blue balls, draw n balls from $N_1 + N_2$ balls, then look at the probability that there are x red balls in the selected n balls.
Mean: $\mu = \mathbb{E}(X) = \sum_{x \in S} x f(x)$.
Variance: $\sigma^2 = Var(X) = \mathbb{E}(X - \mu)^2 = \mathbb{E}(X^2) - \mu^2 = \sum_{i=1}^{k} x_i^2 f(x_i) - \mu^2$. **Standard Deviation:** σ.
r-th Moment: $\mathbb{E}(|X|^r) = \sum_{x \in S} |x|^r f(x) < \infty$; is the moment about the origin.
r-th Moment about b: $\mathbb{E}((X-b)^r) = \sum_{x \in S} (x-b)^r f(x)$; is the moment about b. Facts: μ is the first moment of X about the origin. σ^2 is the second moment of X about μ.
Example of Variance Properties: $Var(aX + b) = a^2 \cdot Var(X)$.
Bernoulli Distribution: A random experiment is called a set of Bernoulli trials if each trial has only two outcomes, has a constant p, and each trial is independent. $f(x) = p$ if $x = 1$, $1 - p$ if $x = 0$, with $0 \leq p \leq 1$. $\mathbb{E}(X) = p$ and $Var(X) = p(1-p)$.
Binomial Distribution: Let X be the number of successes in n independent Bernoulli trials with p. Then $X \sim b(n,p)$. $f(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$. $\mathbb{E}(X) = np$ and $Var(X) = np(1-p)$.
A note: binomial and hypergeometric are similar, but binomial has replacement and one category, and hypergeometric has two categories and no replacement.
Cumulative Distribution Function: $F(x) = \mathbb{P}(X \leq x), x \in (-\infty, \infty)$. For discrete random variables: $f(x) = \mathbb{P}(X = x) = \mathbb{P}(X \leq x) - \mathbb{P}(X \leq x - 1) = F(x) - F(x-1)$.
Geometric Distribution: $f(x) = p(1-p)^{x-1}, x = 1, 2, 3, \ldots$. X represents the draw in which the first success is drawn. $f(x)$ is the probability of getting a success in the x-th draw.
Negative Binomial Distribution: $f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$ and $x = r, r+1, r+2, \ldots$. X is the number of trials it takes to get r successes. $f(x)$ is the probability that the r-th success occurs on the x-th trial.

1.2 Midterm 2 Material

Moment Generating Function: $M(t) = \mathbb{E}(e^{tX}) = \sum_{x \in S} e^{tX} f(x)$ if $f(x)$ is the p.d.f. of some distribution and $t \in V_h(0)$ is finite. *Theorem:* $\mathbb{E}(X^r) = M^{(r)}(0)$, so $\mu = \mathbb{E}(X) = M'(0)$ and $\sigma^2 = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = M''(0) - [M'(0)]^2$. To calculate the m.g.f. and p.d.f. of some random variable given only its moments, use the Taylor series expansion centered at zero. $M(t) = M(0) + M'(0)\left(\frac{t}{1!}\right) + M''(0)\left(\frac{t^2}{2!}\right) + \cdots = 1 + \mathbb{E}(X)\left(\frac{t}{1!}\right) + \mathbb{E}(X^2)\left(\frac{t^2}{2!}\right) + \ldots$.
Poisson distribution: *Definition:* Poisson process counts the number of events occurring in a fixed time/space given a rate λ. Let X_t be the number events which occur in t unit time intervals. $\mathbb{P}(X_t = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$. $\mu = \lambda, \sigma^2 = \lambda$.
Poisson Approximation of Binomial Distribution: If $X \sim b(n,p)$ and n is large while p is small, then X can be approximated as $\hat{X} \sim poi(\lambda)\, s.t.\, \lambda = np$.
Mean: $\mu = \int_{-\infty}^{\infty} x f(x) dx$.
Variance: $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$.
M.G.F.: $M(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx, t \in (-h, h)$.
Percentile: $p = \int_{-\infty}^{\pi_p} f(x) dx = F(\pi_p)$.
Uniform Distribution: $f(x) = \frac{1}{b-a}, a \leq x \leq b$. If $X \sim U(a,b)$, then $\mathbb{E}(X) = \frac{a+b}{2}$, $\sigma^2 = \frac{(b-a)^2}{12}$, and $M(t) = \mathbb{E}(e^{tX}) = \frac{e^{tb} - e^{ta}}{t(b-a)}, t \neq 0$ and $M(0) = 1$.
Exponential Distribution: This describes the waiting time between events in a Poisson process with rate λ. Let $\theta = \frac{1}{\lambda}$. Then if $X \sim Exp(\theta)$, $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x \geq 0$ and 0 otherwise. $\mu = \theta$, $\sigma^2 = \theta^2$, and $M(t) = (1 - \theta t)^{-1}, t < \frac{1}{\theta}$.
Memoryless Property of the Exponential Distribution: What happened in the past does not matter now. Only the present can determine the future. $\mathbb{P}(X > a + b \,|\, X > a) = \mathbb{P}(X > b), \forall a, b \geq 0$.
Gamma Function: $\Gamma(x) = \int_0^{\infty} y^{x-1} e^{-y} dy$, $x > 0$.
Gamma Distribution: $f(t)$ represents the waiting time unitl the α-th occurrence of some Poisson process with rate λ. $f(t) = \frac{\lambda^\alpha t^{\alpha-1}}{(\alpha-1)!} e^{-\lambda t}, t \geq 0$. Then let $\theta = \lambda^{-1}$, then $f(t) = \frac{1}{\Gamma(\alpha)} \theta^\alpha t^{\alpha-1} e^{\frac{-t}{\theta}}$, $t \geq 0$, $\alpha \in \mathbb{R}$. $\mu = \alpha\theta$, $\sigma^2 = \alpha\theta^2$, and $M(t) = (1 - \theta t)^{-\alpha}$. If $\alpha = 1$, gamma$(1, \theta) = \text{Exp}(\theta)$.
χ^2 Distribution: If $X \sim$ gamma(α, θ), and $\theta = 2$ and $\alpha = \frac{r}{2}$, where r is a positive integer, then X is a χ^2 distribution with degree of freedom r. $X \sim \chi^2(r)$. $\mu = r$, $\sigma^2 = 2r$, and $M(t) = (1 - 2t)^{\frac{-r}{2}}$, $t < \frac{1}{2}$. $e^2 = \chi^2(2)$.
Normal Distribution: Bell curve! $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$. $X \sim N(\mu, \sigma^2)$. $X \sim N(0,1)$ is the standard normal distribu-

tion. $\mu = \mu$, $\sigma^2 = \sigma^2$, and $M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$. Standardization: $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
Normal Square Distribution: Let $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$. Then the random variable $V = Z^2 \sim \chi^2(1)$.
Cauchy Distribution: (Why do we even need this disctribution?) $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$. Symmetric about zero, so median is zero, but μ is undefined because the tail of the p.d.f. is too heavy (i.e. each integral of the distribution does not converge). $c.d.f. = F(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$, $x \in \mathbb{R}$.
Joint Probability Density Function: $\mathbb{P}((X,Y) \in A) = \int\int f(x,y)dxdy$.
Marginal Probability Density Function: $f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy$ and the other way around for y.
Mathematical Expectation: $\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$.
Independent Random Variables: Two events A and B are independent iff $f(x,y) = f_x(x) \cdot f_y(y) \, \forall x,y$. This works for both the p.d.f. and the c.d.f.
Trinomial Distribution: This is an extension of the binomial distribution into two dimensions. $f(x_1,x_2) = \mathbb{P}(X_1 = x_1, X_2 = x_2) = \frac{n!}{x_1!x_2!(n-x_1-x_2)!}p_1^{x_1}p_2^{x_2}(1-p_1-p_2)^{n-x_1-x_2}$ for $x_1, x_2 \geq 0$, $x_1 + x_2 \leq n$.
Covariance: $Cov(X,Y) = \mathbb{E}[(X-\mu_x)(Y-\mu_y)], \mu_x = \mathbb{E}(X), \mu_y = \mathbb{E}(Y)$. Useful identity: $Cov(X,Y) = \mathbb{E}(XY) - \mathbb{X}\mathbb{E}(Y)$.
Properties of Covariance: Zero-angle: $Cov(X,X) = Var(X)$. Symmetry: $Cov(X,Y) = Cov(Y,X)$. $Cov(aX+b,Y) = a \cdot Cov(X,Y)$. $Cov(X+Y,W) = Cov(X,W) + Cov(Y,W)$. $Cov(aX+bY, cX+dY) = (ac) \cdot Var(X) + (bd) \cdot Var(Y) + (ad+bc) \cdot Cov(X,Y)$. $Var(aX+bY) = a^2 \cdot Var(X) + 2ab \cdot Cov(X,Y) + b^2 \cdot Var(Y)$.
Correlation Coefficient: $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}$. Uncorrelated: $\rho_{XY} = 0 \Leftrightarrow \sigma_{XY} = 0$. If X and Y are independent, then $\rho_{XY} = 0$, but the implication **does not** go the other way.
I.I.D.: If every marginal probability density function is equal and independent, then all of the random variables are independent and identically distributed (i.i.d.).
Combinations: If Y is a linear combination of independent random variables with constant multiples, then the expectation of Y is simply the sum of the expectations of random variables. The variance of Y is the sum of the variances of the random variables with their constant multiples squared. The moment of a linear combination of random variables is the product of their individual moments.
Summary of Combinations: Suppose X and Y are independent. Then,

- $bern(p) + bern(p) = b(2,p)$
- $b(n_1,p) + b(n_2,p) = b(n_1+n_2,p)$
- $geo(p) + geo(p) = negbin(2,p)$
- $negbin(r_1,p) + negbin(r_2,p) = negbin(r_1+r_2,p)$
- $poi(\lambda_1) + poi(\lambda_2) = poi(\lambda_1+\lambda_2)$
- $exp(\theta) + exp(\theta) = gamma(2,\theta)$
- $gamma(\alpha_1,\theta) + gamma(\alpha_2,\theta) = gamma(\alpha_1+\alpha_2,\theta)$
- $N(\mu_1,\sigma_1^2) + N(\mu_2,\sigma_2^2) = N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

1.3 Final Material

Everything taught in the class is on the final.
Chebyshev's Inequality: If $k \geq 1$: $\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{\sigma^2}{\epsilon^2}$. Let $\epsilon = k\sigma$: $\mathbb{P}(|X-\mu| \geq \epsilon) \leq \frac{1}{k^2}$.
Central Limit Theorem (CLT): $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \to Z$, as $n \to \infty$.
Normal Approximation to Binomial: If $np \geq 5$ and $n(1-p) \geq 5$, then

	$b(n,p)$	$N(\mu,\sigma^2)$
Mean	np	μ
Variance	$np(1-p)$	σ^2

Normal Approximation to Poisson:

	Poisson(λ)	$N(\mu,\sigma^2)$
Mean	λ	μ
Variance	λ	σ^2

Jensen's Inequality: If $g: \mathbb{R} \to \mathbb{R}$ is a convex function and X is a random variable, then $\mathbb{E}[g(X)] \geq g(\mathbb{E}(X))$.
Maximum Likelihood Estimator: Maximize the likelihood function subject to the parameter space.

1. Construct likelihood function, $F(x;\theta) = \Pi_{i=1}^n(f(x_i))$.
2. If applicable, take the log of this to make things easier to differentiate and to change the product into a sum. $l(x;\theta)$.
3. Differentiate to get $\frac{dl(x;\theta)}{d\theta}$.
4. Assume convexity (ha) and find optimal θ by solving $\frac{dl(x;\theta)}{d\theta} = 0$.

Sample Mean: $\frac{1}{n}\sum_{i=1}^n X_i \equiv \overline{X}$.
Sample Variance: For z distribution: $\frac{1}{n}\sum_{i=1}^n(X_i - \overline{X})$, and for t distribution: $\frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X})$.
Method of Moments Estimator: Construct a system of equations using moments and sample data. For example, equate theoretical mean with sample mean, and theoretical variance with sample variance, then solve. `<><><maybe put an example here<><>><`
Confidence Intervals for Means:

	Known Variance	Unknown Variance
Two Side	$\left[\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$	$\left[\overline{X} - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right]$
Lower	$\left[\overline{X} - z_\alpha\frac{\sigma}{\sqrt{n}}, \infty\right)$	you can
Upper	$\left(-\infty, \overline{X} + z_\alpha\frac{\sigma}{\sqrt{n}}\right]$	figure it out

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