STAT 630–Formulas for the Final Exam

Cumulative Distribution Function
The cdf of a random variable X is a function $F_X(x) = P(X \leq x)$ for each x.

Relationship of CDF and PDF for a Continuous RV $f_X(x) = \frac{d}{dx} F_X(x)$.

Function of a Discrete RV
Let $Y = h(X)$ where X is a discrete rv with pmf $p_X(x)$. Then the pmf of Y is $p_Y(y) = \sum_{\{x:h(x)=y\}} p_X(x)$.

Function of a Continuous RV
Let $Y = h(X)$ where X is a continuous rv with pdf $f_X(x)$. Then the cdf of Y is

$$F_Y(y) = P[h(X) \leq y] = \int_{\{x:h(x) \leq y\}} f_X(x) dx$$

If h is differentiable and strictly monotonic on some interval I which includes the range of X, the pdf of Y equals

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right|.$$

Joint Probability Mass Function $p(x, y) = P(X = x, Y = y)$

Joint Probability Density Function A joint pdf f is a nonnegative function such that

$$P\left((X, Y) \in A\right) = \int_A \int f(x, y)\, dx dy.$$

Bivariate Distribution Function This is the function F such that $F(x, y) = P(X \leq x, Y \leq y)$.

Obtaining Joint PDF from CDF If X and Y are continuous rvs, $f(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$

Marginal Distributions When X and Y have joint pmf p or joint pdf f, the marginal pmf or pdf of X is

$$p_X(x) = \sum_y p(x, y), \text{ for } X,\, Y \text{ discrete} \qquad f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \text{for } X,\, Y \text{ continuous.}$$

Conditional PMF or PDF When X and Y have joint pmf p or pdf f and X has marginal pmf p_X or pdf f_X, the conditional pmf or pdf of Y given that $X = x$ is

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)}, \qquad \text{or} \qquad f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}.$$

Independent RVs When X and Y have joint pmf p or pdf f, they are independent iff

$$p(x, y) = p_X(x) p_Y(y) \qquad \text{or} \qquad f(x, y) = f_X(x) f_Y(y), \text{ all } x, y.$$

Convolutions When X and Y are independent continuous rvs with pdfs f_X and f_Y, $Z = X + Y$ has cdf and pdf

$$F_Z(z) = \int_{-\infty}^{\infty} f_Y(y) F_X(z - y) dy \qquad f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx.$$

When X and Y are independent discrete rvs with pmfs p_X and p_Y, the pmf of $Z = X + Y$ is

$$p_Z(z) = \sum_{\{(x,y):x+y=z\}} p_X(x) p_Y(y).$$

Maxima and Minima: When $X_1, \ldots, X_n$ are continuous rvs with the same cdf F, the cdfs of the maximum U and minimum V are, respectively, $F_U(u) = [F(u)]^n$ and $F_V(v) = 1 - (1 - F(v))^n$.

Expectations: Let p or f be the pmf or pdf of a random variable X. The expectation of $h(X)$ is

$$E[h(X)] = \begin{cases} \int_{-\infty}^{\infty} h(x)f(x)\,dx, & X \text{ continuous,} \\ \sum_x h(x)p(x), & X \text{ discrete.} \end{cases}$$

Expectation of a linear function: $E(aX + b) = aE(X) + b$

Expectation of a sum: $E(b + a_1X_1 + \cdots + a_nX_n) = b + a_1E(X_1) + \cdots + a_nE(X_n)$

Expectation of a product of independent rvs: Suppose $X_1, \ldots, X_n$ are independent rvs. Then

$$E\left(\prod_{i=1}^{n} h(X_i)\right) = \prod_{i=1}^{n} E(h(X_i))$$

Variance: Let $\mu = E(X)$. The variance of X is $\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - (E(X))^2$.

Covariance and correlation:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y), \qquad \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Covariance of sums and variance of a sum:

$$\text{Cov}(a + \sum_{i=1}^{n} b_iX_i, c + \sum_{j=1}^{m} d_jY_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_id_j\text{Cov}(X_i, Y_j), \ \text{Var}\Big(b + \sum_{i=1}^{n} a_iX_i\Big) = \sum_{i=1}^{n} a_i^2\text{Var}(X_i) + \sum_{i=1}^{n}\sum_{j\neq i} a_ia_j\text{Cov}(X_i, X_j)$$

Moments: The moments of X are $E(X^k)$, $k = 1, 2, \ldots$.

Moment generating function: The mgf of X is $M(s) = E(e^{sX})$.

$$E(X^k) = \left.\frac{d^kM(s)}{ds^k}\right|_{s=0}, \ M_{aX+b}(s) = e^{bs}M_X(as), \ M_{X_1+\cdots+X_n}(s) = \prod_{i=1}^{n} M_{X_i}(s), \text{ if } X_1, \ldots, X_n \text{ are independent}$$

Conditional mean: The conditional mean of Y given $X = x$ is

$$E(Y|x) = \int_{-\infty}^{\infty} yf_{Y|X}(y|x)dy \quad \text{or} \quad E(Y|x) = \sum_y yp_{Y|X}(y|x)$$

$$E[E[Y|X]] = E[Y]$$

Conditional variance: The conditional variance of Y given $X = x$ is

$$\text{Var}(Y|X = x) = E[(Y - E(Y|x))^2|X = x]$$

$$\text{Var}(Y) = \text{Var}[E(Y|X)] + E[\text{Var}(Y|X)]$$

Markov Inequality: For any nonnegative rv X and $a > 0$, $P[X \geq a] \leq \frac{E(X)}{a}$

Chebyshev Inequality: For all $a > 0$, $P(|X - \mu_X| \geq a) \leq \frac{\sigma^2}{a^2}$

Convergence in probability: The sequence $Z_1, Z_2, \ldots$ converges in probability to the constant b if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|Z_n - b| \geq \varepsilon) = 0.$$

Law of Large Numbers: If $X_1, X_2, ...$ is a sequence of independent rvs with mean μ and variance σ^2, then $\bar{X}_n$ converges in probability to μ.

Convergence in distribution: The sequence of rvs $X_1, X_2, \ldots$ converges in distribution to the rv X if for all x where $F_X(x)$ is continuous, $\lim_{n\to\infty} P(X_n \le x) = P(X \le x) = F_X(x)$.
Central Limit Theorem: If $X_1, X_2, ..., X_n$ is a random sample from a distribution with mean μ and variance σ^2, $S_n = \sum_{i=1}^n X_i$, and $\bar{X}_n = S_n/n$, then

$$\lim_{n\to\infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le z\right) = \lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z\right) = \Phi(z).$$

Chi-Squared distribution: The chi-squared distribution with n degrees of freedom (df) is a gamma distribution with $\alpha = n/2$ and $\lambda = 1/2$. If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2_1$. If $X_1, \ldots, X_k$ are independent chi-squared rvs with dfs $n_1, \ldots, n_k$, respectively, then $X_1 + \cdots + X_k \sim \chi^2_{n_1+\cdots+n_k}$.
F **distribution:** If X and Y are independent chi-squared rvs with m and n degrees of freedom, respectively, the rv

$$F = \frac{X/m}{Y/n}, \qquad \text{with pdf} \quad f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1}\left(1+\frac{m}{n}w\right)^{-(m+n)/2}, \ w \ge 0$$

has an F distribution with m, n degrees of freedom.
t **distribution:** If $Z \sim N(0,1)$ and $Y \sim \chi^2_n$ are independent rvs, then the rv

$$T = \frac{Z}{\left(\frac{Y}{n}\right)^{1/2}} \quad \text{with pdf} \quad f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}, \ -\infty < t < \infty$$

has a t distribution with n df.
Sampling from the normal distribution: For a random sample from a $N(\mu, \sigma^2)$ distribution, the sample mean $\bar{X}$ and the sample variance $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ are independent random variables satisfying

$$\bar{X} \sim N(\mu, \sigma^2/n) \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Maximum likelihood estimation: The likelihood function for a random sample $X_1, \ldots, X_n$ from a distribution with pmf/pdf $f_\theta(x)$ is $L(\theta|x_1, \ldots, x_n) = f_\theta(x_1) \cdot f_\theta(x_2) \cdots f_\theta(x_n)$. The maximum likelihood estimation $\hat{\theta}$ maximizes the likelihood, or equivalently, maximizes the log-likelihood by solving the score equation:

$$\frac{\partial}{\partial\theta} \log\left[L(\theta|x_1, \ldots, x_n)\right] = 0 \ \text{ for } \theta.$$

Invariance property of MLE: If $\hat{\theta}$ is the mle of θ and $\psi(\theta)$ is a one-to-one function, then $\psi(\hat{\theta})$ is the mle of $\psi(\theta)$.
Method of moment estimators: Equate the k^{th} *population* moment $E_{\theta_1,\ldots,\theta_m}[X^k]$ with the k^{th} *sample* moment $m_k = \frac{1}{n}\sum x_i^k$, for $k = 1, \ldots, m$ and solve for $\theta_1, \ldots, \theta_m$.
Bias: The bias of an estimator T of $\psi(\theta)$ is $\text{Bias}_\theta(T) = E_\theta(T) - \psi(\theta)$.
Standard error: The standard error of an estimator T is $\sqrt{\text{Var}_\theta(T)}$.
Mean squared error: The MSE of an estimator T of $\psi(\theta)$ is $\text{MSE}_\theta(T) = E_\theta[(T - \psi(\theta))^2] = \text{Var}_\theta(T) + \text{Bias}_\theta(T)^2$.
Consistency: A sequence $\{Tn\}$ of estimators of $\psi(\theta)$ is consistent if $T_n \xrightarrow{P} \psi(\theta)$.
Fisher information: Let X be a rv with pmf or pdf $f(x|\theta)$. Define the log-likelihood for one observation: $\log f_\theta(x)$. The Fisher information in the rv X is defined as

$$I(\theta) = E_\theta\left[\left(\frac{\partial \log f_\theta(X)}{\partial\theta}\right)^2\right] = -E_\theta\left(\frac{\partial^2 \log f_\theta(X)}{\partial\theta^2}\right).$$

The Fisher information $I_n(\theta)$ in a random sample $X_1, \ldots, X_n$ from $f_\theta(x)$ is $I_n(\theta) = nI(\theta)$.

Asymptotic normality of the MLE: Suppose that $X_1, \ldots, X_n$ form a random sample from a population with pf or pdf $f_\theta(x)$. Let $I(\theta)$ be the Fisher information in a single observation. Let $\hat{\theta}$ denote the mle of θ. The MLE is asymptotically normal:

$$\frac{\sqrt{n}(\hat{\theta}-\theta)}{\sqrt{I(\theta)^{-1}}} \xrightarrow{d} N(0,1) \quad \text{as} \quad n \longrightarrow \infty \qquad \text{or} \qquad \hat{\theta} \sim \text{Asymp}N\left(\theta, \frac{1}{nI(\theta)}\right)$$

Confidence interval: Let $X_1, \ldots, X_n$ form a random sample from a distribution with parameter θ. Suppose that $l(X_1, \ldots, X_n)$ and $u(X_1, \ldots, X_n)$ are statistics such that $P[l(X_1, \ldots, X_n) < \psi(\theta) < u(X_1, \ldots, X_n)] = \gamma$, for $0 < \gamma < 1$ and any $\theta \in \Theta$. We observe the data $X_1 = x_1, \ldots, X_n = x_n$ and compute $l(x_1, \ldots, x_n)$ and $u(x_1, \ldots, x_n)$. The interval $(l(x_1, \ldots, x_n), u(x_1, \ldots, x_n))$ is a confidence interval for $\psi(\theta)$ with confidence coefficient γ.

Pivot: A pivot, $W(X_1, \ldots, X_n, \theta)$, is a random variable whose distribution does not depend on θ. Suppose that $P(c_1 < W(X_1, \ldots, X_n, \theta) < c_2) = \gamma$. A level γ confidence interval for θ is
$C(x_1, \ldots, x_n) = \{\theta : c_1 < W(x_1, \ldots, x_n, \theta) < c_2\}$.

Approximate level γ confidence interval based on MLE:

$$\hat{\theta} \pm Z_{(1+\gamma)/2} ASE(\hat{\theta}) \quad \text{where} \quad ASE(\hat{\theta}) = \sqrt{\frac{1}{nI(\hat{\theta})}}$$

Bayes Model: The unknown parameter θ is a random variable with a prior distribution $\pi(\theta)$. For a given value of θ, the data s have a pdf or pmf $f_\theta(s)$.

Posterior Distribution: The conditional distribution of θ given $S = s$ is

$$\pi(\theta|s) = \frac{f(s,\theta)}{m(s)} = \frac{f_\theta(s)\pi(\theta)}{\int f_\theta(s)\pi(\theta)d\theta}.$$

Hypotheses: Statements about the parameter denoted by H_0 (null hypothesis) and H_a (alternative hypothesis).

Type I error: Reject H_0 when H_0 true, α is the probability of a Type I error.

Type II error: Fail to reject H_0 when H_0 false, $1 - \beta$ is the probability of a Type II error.

Power: β=the probabilility of rejecting H_0 when H_0 is false. The power function depends on the value of the parameter specified by H_a.

The **test statistic** is used to make the decision. The **rejection region** is the set of values of the test statistic for which H_0 is rejected.

$P-$**value:** The $P-$value corresponding to the observation $X = x$ is the smallest level of significance at which H_0 can be rejected.

Neyman-Pearson Lemma For testing $H_0 : \theta = \theta_0$ versus $H_a : \theta = \theta_1$, the level α test that rejects H_0 for large values of the **likelihood ratio** $\frac{f_{\theta_1}(x_1,\ldots,x_n)}{f_{\theta_0}(x_1,\ldots,x_n}$ is the most powerful level α test.

The generalized likelihood ratio test for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ rejects H_0 for large values of $LR = \frac{f_{\hat{\theta}}(x_1,\ldots,x_n)}{f_{\theta_0}(x_1,\ldots,x_n)}$.

Asymptotic distribution of the LR statistic: Under the regularity conditions for the asymptotic normality of the MLE, $2\log(LR)$ has approximately a chi-squared distribution with $df = 1$.

Wald and score statistics: The following statistics are asymptotically equivalent to the LR statistic for testing $H_0 : \theta = \theta_0$:

$$W = \frac{(\hat{\theta}-\theta_0)^2}{I_n^{-1}(\hat{\theta})} \qquad \text{and} \qquad S = \frac{U(\theta_0)^2}{I_n(\theta_0)} \qquad \text{where} \qquad U(\theta) = \frac{\partial \log(f_\theta(x_1, \ldots, x_n))}{\partial\theta}$$

Equivalence of tests and confidence intervals:

Suppose that $(l(x_1, \ldots, x_n), u(x_1, \ldots, x_n))$ is a level $1 - \alpha$ confidence interval for θ. Then a level α test of $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$ rejects H_0 if θ_0 is not inside the confidence interval.

Binomial distribution:

$$p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x = 0, 1, \ldots, n, \quad \text{Mean} = n\theta, \quad \text{Variance} = n\theta(1-\theta), \quad M(s) = (1-\theta+\theta e^s)^n$$

Binomial sum and geometric sum

$$\sum_{x=0}^{n}\binom{n}{x}a^x b^{n-x} = (a+b)^n, \qquad \sum_{x=0}^{\infty}\alpha^x = \frac{1}{1-\alpha}, \text{ for } 0 < \alpha < 1$$

Hypergeometric distribution

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad \max\{0, n-(N-M)\} \leq x \leq \min(n, M), \text{ Mean} = n\left(\frac{M}{N}\right)$$

Negative binomial distribution

$$p(x) = \binom{r-1+x}{x}\theta^r(1-\theta)^x, \; x = 0, 1, 2, \ldots \text{ Mean} = \frac{r(1-\theta)}{\theta}, \text{ Var} = \frac{r(1-\theta)}{\theta^2}, \; M(s) = \left(\frac{\theta}{1-(1-\theta)e^s}\right)^r$$

Geometric distribution (Negative binomial with $r = 1$)

$$p(x) = \theta(1-\theta)^x, \quad x = 0, 1, 2, \ldots \qquad \text{Mean} = \frac{1-\theta}{\theta}$$

Poisson distribution

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0, \qquad \text{Mean} = \lambda \qquad \text{Variance} = \lambda, \quad M(s) = e^{\lambda(e^s-1)}$$

Normal distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < \mu < \infty, \sigma > 0$$

$$\text{Mean} = \mu \qquad \text{Variance} = \sigma^2, \qquad M(s) = \exp(\mu s + \sigma^2 s^2/2)$$

Gamma distribution

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}I_{(0,\infty)}(x), \text{ where } \alpha > 0 \text{ and } \lambda > 0.$$

$$\text{Mean} = \frac{\alpha}{\lambda}, \qquad \text{Variance} = \frac{\alpha}{\lambda^2}, \qquad M(t) = \left(\frac{\lambda}{\lambda-t}\right)^\alpha, \text{ for } t < \lambda.$$

Exponential distribution (Gamma with $\alpha = 1$)

$$f(x) = \lambda e^{-\lambda x}I_{(0,\infty)}(x), \quad \lambda > 0, \qquad \text{Mean} = \frac{1}{\lambda}, \qquad \text{Variance} = \frac{1}{\lambda^2}$$

Uniform distribution

$$f(x) = \frac{1}{(R-L)}\cdot I_{(L,R)}(x), \quad -\infty < L < R < \infty, \qquad \text{Mean} = \frac{(L+R)}{2} \qquad \text{Variance} = \frac{(R-L)^2}{12}$$

Beta distribution

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, \; 0 \leq x \leq 1, \quad \text{Mean} = \frac{a}{a+b}, \quad \text{Variance} = \frac{ab}{(a+b)^2(a+b+1)}.$$

A Few Indefinite Integrals and One Definite Integral

$$\int x^n dx = \frac{x^{n+1}}{n+1} \qquad \text{except for } n = -1$$

$$\int \frac{1}{x}dx = \log_e(x)$$

$$\int u dv = uv - \int v du \qquad \text{integration by parts}$$

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx, \qquad a > 0, \qquad \Gamma(a+1) = a\Gamma(a), \qquad \Gamma(1/2) = \sqrt{\pi}$$