EXAM 2 FORMULAS

Formulas in gray will not be provided: **memorize these**.

Theorem. *For any random variable X and any constants a and b:*

1. $E(aX+b) = aE(X)+b$ *and*
2. $Var(aX+b) = a^2\,Var(X)$.

Theorem. *Let $X_1, X_2, \dots, X_n$ be any random variables. Then $E(X_1+X_1+\cdots+X_n) = E(X_1)+E(X_2)+\cdots+E(X_n)$.*

Theorem. *Let $X_1, X_2, \dots, X_n$ be independent random variables. Then $Var(X_1+X_2+\cdots+X_n) = Var(X_1)+Var(X_2)+\cdots+Var(X_n)$.*

Theorem (Propagation of error formula)**.** *If X is a random variable with mean μ_X and standard deviation σ_X and $g(x)$ is a differentiable function, then*

1. $E[g(X)] \approx g(\mu_X)$
2. $Var[g(X)] \approx [g'(\mu_X)\sigma_X]^2$

Definition. A **random sample of size** n is a set of independent identically distributed (iid) random variables $X_1, X_2, \dots X_n$. Some **sample statistics**:

1. The **sample total** $T = \sum_{i=1}^n X_i$
2. The **sample mean**: $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$
3. The **sample variance**: $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (X_i - \overline{X})^2$

Theorem. *For any random sample from a population with mean μ and variance σ^2:*

1. $E(T) = n\mu$ *and* $Var(T) = n\sigma^2$
2. $E(\overline{X}) = \mu$ *and* $Var(\overline{X}) = \frac{\sigma^2}{n}$
3. $E(S^2) = \sigma^2$

Definition. A sample statistic $\hat{\Theta}$ is an **unbiased estimator** of population parameter θ if $E(\hat{\Theta}) = \theta$.

Theorem (Central Limit Theorem)**.** *If $\overline{X}$ is a the mean of a large random sample from any population, then $\overline{X}$ is approximately normally distributed (with mean and variance given in the theorem above).*

Definition. Let X and Y be jointly distributed discrete RVs with joint PMF $f(x,y) = P(X=x, Y=y)$.

1. The **marginal PMF** of X is $p_X(x) = P(X=x) = \sum_y p(x,y)$.
2. X and Y are **independent** if $f(x,y) = f_X(x)f_Y(y)$.
3. The **covariance** of X and Y is $\mathrm{Cov}(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y)$.
4. **Pearson's correlation coefficient** is $\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$.

Definition. Confidence intervals:

1. z_β is the z-critical value: $P(Z > z_\beta) = \beta$:

β	0.1	0.05	0.025	0.01	0.005
z_β	1.281552	1.644854	1.959964	2.326348	2.575829

2. If $\overline{x}$ is the mean of a random sample of size n (with n large) from a population with mean μ and standard deviation σ, then a $100(1-\alpha)\%$ confidence interval for μ is $\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$.
3. If p is the proportion of a random sample of size n (with n large) from a population having a Bernoulli distribution with parameter θ, then an approximate $100(1-\alpha)\%$ confidence interval for θ is $p \pm z_{\alpha/2}\sqrt{\dfrac{p(1-p)}{n}}$.

Date: March 23, 2024.

4. $t_{\beta,\nu}$ is the t-critical value for a t-distribution with parameter ν: $P(T > t_{\beta,\nu}) = \beta$. Note: t-based confidence intervals won't be on the test, but two snuck on to the WeBWorK.
5. If $\overline{x}$ is the mean of a random sample of size n from a normally-distributed population with mean μ, then a $100(1-\alpha)\%$ confidence interval for μ is $\overline{x} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$.

Definition. The total number of successes in n independent, identically distributed (iid) Bernoulli trials with parameter p is a random variable with a **binomial** distribution. The PMF of a random variable X having a binomial distribution with parameters n and p is $p(x) = \binom{n}{x}p^x(1-p)^{n-x}$ for $x = 0, 1, \ldots, n$. The **R** syntax for the CDF is $P(X \leq x) = \texttt{pbinom}(\texttt{x}, \texttt{n}, \texttt{p})$.

Proposition. *The mean and variance of a binomial distribution are $\mu = np$ and $\sigma^2 = np(1-p)$.*

Definition. The number of trials until the first success in a sequence of independent, identically distributed (iid) Bernoulli trials with probability of success p is a random variable with a **geometric distribution**. The PMF of a random variable X having a geometric distribution with parameter p is

$$P(X = x) = p(1-p)^{x-1} \text{ for } x = 1, 2, 3, \ldots$$

The CDF is $P(X \leq x) = 1 - (1-p)^x$ for $x = 1, 2, 3, \ldots$.

Proposition. *The mean and variance of a geometric distribution are $\mu = \frac{1}{p}$ and $\sigma^2 = \frac{1-p}{p^2}$.*

Definition. A **Poisson** random variable X with parameter $\lambda > 0$ has the PMF $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, 2, \ldots$. The **R** syntax for the CDF is $P(X \leq x) = \texttt{ppois}(\texttt{x}, \lambda)$.

Proposition. *The mean and variance of a Poisson distribution are $\mu = \lambda$ and $\sigma^2 = \lambda$.*

Definition. A random variable X having a **uniform continuous** distribution on the interval $[\alpha, \beta]$ has the PDF: $f(x) = \frac{1}{\beta - \alpha}$ if $\alpha < x < \beta$.

Proposition. *The mean and variance of a uniform continuous distribution on $[\alpha, \beta]$ are $\mu = \frac{\alpha+\beta}{2}$ and $\sigma^2 = \frac{(\beta-\alpha)^2}{12}$.*

Definition. A random variable X having an **exponential** distribution with parameter $\lambda > 0$ has PDF $f(x) = \lambda e^{-\lambda x}$ if $x > 0$ and CDF $P(X \leq x) = 1 - e^{-\lambda x}$ if $x > 0$.

Proposition. *The mean and variance of an exponential distribution with parameter λ are $\mu = \frac{1}{\lambda}$ and variance $\sigma^2 = \frac{1}{\lambda^2}$.*

Definition. A random variable X having a **normal** distribution with mean μ and standard deviation σ has the PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for all $x \in \mathbb{R}$

R Implementation. If $X \sim \text{Normal}(\mu, \sigma)$, then the CDF is `pnorm(x, μ, σ)` (note that R wants the standard deviation, not the variance). The parameters μ and σ are optional; if ommitted, they default to $\mu = 0$ and $\sigma = 1$ (a **standard normal distribution**).
