EXAM 2 FORMULAS

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

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

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**: $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$

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* $Var(\overline{X}) = \frac{\sigma^2}{n}$

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

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 $p(x,y) = P(X=x, Y=y)$.

1. The **marginal PMF** of X is $p_X(x) = P(X=x) = \sum_y p(x,y)$.
2. The **conditional PMF** of Y given $X=x$ is $p_{Y|X=x}(y) = P(Y=y|X=x) = \frac{p(x,y)}{p_X(x)}$.
3. The **regression function** of Y on X is $\mu_{Y|X=x} = \sum_y y p_{Y|X=x}(y)$.
4. X and Y are **independent** if $p(x,y) = p_X(x)p_Y(y)$.
5. The **covariance** of X and Y is $\text{Cov}(X,Y) = \sigma_{X,Y} = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E(XY) - E(X)E(Y)$.
6. **Pearson's correlation coefficient** is $\rho = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$.

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 $x = 0, 1, \dots, n$.

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

R Implementation. If $X \sim \text{binom}(n,p)$, then the CDF is `pbinom(x, n, p)`.

Definition. If independent, identically distributed Bernoulli trials with parameter p are repeated until the first success, then the total number of trials (counting the success) has a **geometric** distribution. The PMF is $p(x) = (1-p)^{x-1}p$ for $x = 1, 2, 3, \dots$ and the CDF is $F(x) = 1-(1-p)^x$ for $x = 1, 2, 3, \dots$.

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 with parameter $\lambda > 0$ has the PMF $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, 2, \dots$.

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

R Implementation. If $X \sim \text{pois}(\lambda)$, then the CDF is `ppois(x, λ)`.

Definition. A random variable X with 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 $\sigma^2 = \frac{(\beta-\alpha)^2}{12}$.*

Date: March 21, 2023.

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

Proposition. *An exponential distribution with parameter λ has mean $\mu = \frac{1}{\lambda}$ and variance $\sigma^2 = \frac{1}{\lambda^2}$.*

R Implementation. If $X \sim \exp(\lambda)$, then the CDF is `pexp(x, λ)`.

Definition. A random variable with a **normal** distribution with parameters μ and $\sigma^2 > 0$ has the PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for all $x \in \mathbb{R}$

Proposition. *A normal distribution with parameters μ and σ^2 has mean $\mu = \mu$ and variance $\sigma^2 = \sigma^2$.*

R Implementation. If $X \sim \mathrm{N}(\mu, \sigma^2)$, 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**).