Chapter 5 Some Discrete Probability Distributions

Binomial: Two possible outcomes from each trial.
Binomial Distribution: $b(x;n,p) = \binom{n}{x} p^x q^{n-x}$ x: num of successes. n: num of indep. trials. p: prob. of success. q: prob. of failure. $\mu = np$, $\sigma^2 = npq$
Multinomial: $f(x_1, x_2, ...x_k; p_1, p_2, ...p_k, n) = \binom{n}{x_1, x_2 ... x_k} p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$. $n = \sum_{i=1}^k x_i$, and $\sum_{i=1}^k p_i = 1$
Hypergeometric: Choosing successful items.
Hypergeometric Distribution: $h(x;N,n,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$. $max(0, n-(N-k)) \leq x \leq min(n,k)$. x: num of successes. N: num of items. n: num of selection. k: num of total successes. $\mu = \frac{nk}{N}$, $\sigma^2 = \frac{N-n}{N-1} n \frac{k}{N}(1 - \frac{k}{N})$
Estimating Hypergeometric using Binomial: If n is small compared to N: $(n/N) \leq 0.05$.
Multivariate: $f(x_1, x_2, ...x_k; a_1, a_2, ...a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}}$. $n = \sum_{i=1}^k x_i$, $N = \sum_{i=1}^k a_i$.
Negative Binomial Distribution: Prob. the kth success will happen by the xth trial. $b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}$. x: trial number. k: success number. p: prob. success. q: prob. failure.
Geometric Distribution: Prob. the xth trial is the first success. $g(x;p) = pq^{x-1}$. x: trial number. p: prob. success. q: prob. failure. $\mu = \frac{1}{p}$. $\sigma^2 = \frac{1-p}{p^2}$.
Poisson Distribution: Prob. something happens x times in t time. $p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$. x: num of times. λ: average number of outcomes per time period. t: time interval.
newline

Chapter 6 Some Continuous Probability Distributions

Uniform Distribution: Equal Probability throughout interval. $f(x;A,B) = \frac{1}{B-A}$ if $A \leq x \leq B$, 0 otherwise. $\mu = \frac{A+B}{2}$, $\sigma^2 = \frac{(B-A)^2}{12}$.
Normal Distribution: Bell curve. $n(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. x: select time. μ: mean. σ: standard deviation.
Standard Normal: A normal distribution where mean is 0 and variance is 1. $Z = \frac{X-\mu}{\sigma}$
Estimating Binomial with Normal: For large n. $P(X \leq x) \approx P(Z \leq \frac{x+0.5-np}{\sqrt{npq}})$
Gamma Function: $\Gamma(n) = (n-1)!$. $\Gamma(1) = 1$. $\Gamma(1/2) = \sqrt{\pi}$.
Gamma Distribution: Wait time, reliability. $f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$, $x > 0$. or 0 otherwise. $\mu = \alpha\beta$, $\sigma^2 = \alpha\beta^2$.
Exponential Distribution: Special case of Gamma where $\alpha = 1$. $f(x;\beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$ where $x > 0$. 0 elsewhere. β: mean time between failures. α: number of events. $\mu = \beta$, $\sigma^2 = \beta^2$.
Chi-Squared Distribution: Special case of Gamma where $\alpha = v/2$ and $\beta = 2$. $f(x;v) = \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2}$, $x > 0$. 0 elsewhere. v: degrees of freedom. $\mu = v$, $\sigma^2 = 2v$.
Beta Function: $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, $\alpha, \beta > 0$.
Beta Distribution: $f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$. 0 elsewhere. $\mu = \frac{\alpha}{\alpha+\beta}$, $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.
Lognormal Distribution: if $ln(X)$ is a normal distribution. $f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-1/2\sigma^2(ln(x)-\mu)^2}$, $x \geq 0$. 0 if $x < 0$. mean $= e^{\mu+\sigma^2/2}$, variance $= e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

Chapter 8 Fundamental Sampling Distributions and Data Descriptions

Central Limit Theorem: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ as $n \to \infty$ is a standard normal distribution. $\bar{X}$: mean of random sample size. μ: mean of population. σ: standard deviation of population. n: sample size.
Difference of Means: Two populations, samples, means, and variances. $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$. is approx. a standard normal variable. $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$. $\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
Finding Chi-Squared from Variance: $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \frac{(x_1 - \bar{x})^2}{\sigma^2}$. Degrees of freedom is $v = n-1$, n is sample size.
t-Distribution: $T = \frac{Z}{\sqrt{V/v}}$ or $T = \frac{\bar{x}-\mu}{S/\sqrt{n}}$. then $h(t) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi v}}(1 + \frac{t^2}{v})$ from $-\infty < t < \infty$. Z: standard normal RV. V: chi2 RV. v: degrees of freedom.
F-Distribution: $h(f) = \frac{\Gamma((v_1+v_2)/2)(v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \cdot \frac{f^{(v_1/2)-1}}{(1+v_1 f/v_2)^{(v_1+v_2)/2}}$. for $f > 0$, 0 if $f \leq 0$. $F = \frac{U/v_1}{V/v_2}$. V,U: indep. RV with chi2 distribution. v_1, v_2: degrees of freedom. $f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}$. $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$
Normal Q-Q Plot: Set of observations for normal distribution, will be straight if normal. 1) order data ascending. 2) split normal distribution to $n+1$ parts. 3) match the data to the distribution x=data, y=normal. Match smallest with smallest.

Chapter 9 One- and Two-Sample Estimation Problems

CI on μ, σ^2 known: $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. $100(1-\alpha)\%$ confidence interval. $z_{\alpha/2}$ is the z-value leaving the area of $\alpha/2$ to the right. $100(1-\alpha)\%$ confident that error will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Also confident error will not exceed size e as $n = (\frac{z_{\alpha/2}\sigma}{e})^2$. One sided bound? just take one side of the equation, + is upper.
CI on μ, σ^2 unknown: $\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$ where $t_{\alpha/2}$ is the t-value with $v = n-1$ degress of freedon leaving $\alpha/2$ area to the right.
Confidence limits on μ, σ^2 unknown: $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$. standard error is $\frac{\sigma}{\sqrt{n}}$
CI for $\mu_1 - \mu_2$, σ_1^2 and σ_2^2 known: $(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.
Pooled Estimate of Variance: $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$
CI for $\mu_1 - \mu_2$, $\sigma_1^2 = \sigma_2^2$ but both unknown: $(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
CI for $\mu_1 - \mu_2$, $\sigma_1^2 \neq \sigma_2^2$ and both unknown: $(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
CI for $\mu_D = \mu_1 - \mu_2$ for Paired Observations: For the table, find d_i which is the difference between items. $\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}}$