University of Waterloo
Statistics

Brent Scheibelhut

Compiled below are a list of useful properties, general formula manipulation, etc.

1 General

a) $\frac{d}{dy}\sum f(x) = \sum \frac{d}{dy} f(x)$

b) $\prod f(x)^{1/y} = f(x)^{\sum 1/y}$

c)

d)

e)

2 Graphs

a) Tail of a graph is the smaller end that (possibly) can stretch for a while

b) Negatively Skewed = Skewed to the left = Tail on left (small left end) = Mean less than median

c) Positively Skewed = Skewed to the right = Tail on right (small right end) = Mean greater than median

d) Correlation is measure of **linear** dependency

3 Basics

3.1 Expectation(Mean)

a) $E(X) = \sum_{allx} xf(x)$ discrete random variable X, probability function f(x)

b) $E(g(X)) = \sum_{allx} g(x)f(x)$

c) Positively Skewed = Skewed to the right = Tail on right (small right end) = Mean greater than median

d) Correlation is measure of **linear** dependency

3.2 Variance

a)

3.3 Gamma Function

a) $\gamma(\alpha) = (\alpha - 1)!$

b) $\gamma(\alpha) = (\alpha - 1)\gamma(\alpha - 1)$

c) $\gamma(0.5) = \sqrt(\pi)$

d) $\gamma(-1) = \sqrt(2)$

3.4 Ln Function

- $ln(x * y) = ln(x) + ln(y)$
- $ln(x/y) = ln(x) - ln(y)$
- $ln(x^y) = y * ln(x)$
- $f(x) = ln(x) \Rightarrow f'(x) = \frac{1}{x}$
- $\int ln(x)dx = x * (ln(x) - 1) + C$
- $ln(x)$ is undefined when $x \leq 0$
- $ln(1) = 0$
- $\lim_{x \to +\infty} ln(x) = +\infty$
- $\lim_{x \to 0^+} ln(x) = -\infty$

3.5 Definite/Indefinite Integrals

If f(x) and g(x) are defined and continuous on [a, b], except maybe at a finite number of points and c $\in$ [a,b]:

- $\int_a^b (f(x) + g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx$ *same for indefinite
- $\int_a^b (f(x) - g(x))dx = \int_a^b f(x)dx - \int_a^b g(x)dx$ *same for indefinite
- $\int_a^b \alpha f(x)dx = \alpha \int_a^b f(x)dx$ for any arbitrary constant α *same for indefinite
- $\int_c^c f(x)dx = 0$
- $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_b^c f(x)dx$
- $\int_b^a f(x)dx = \int_a^b f(x)dx$
- $\int x^n dx = \frac{1}{n+1} x^{n+1} + C \forall n \neq -1$
- $\int e^x dx = e^x + C$

4 Linear Combinations

4.1 Independence

a) If X and Y are independent than Cov(X,Y)=0

4.2 Linear Combinations

Note $E(X) = \mu$ and $Var(X) = \sigma^2$

a) Suppose random variables X and Y are independent. Then, if $g(X)$ and $f(Y)$ are and two functions. $E[g(X)f(Y)] = E[g(X)]E[f(Y)]$

b) $E(aX + bY) = aE(X) + bE(Y)$, when a and b are constants

c) $E(X + Y) = E(X) + E(Y)$

d) $E(X - Y) = E(X) - E(Y)$

e) $E(\sum a_i X_i) = \sum a_i E(X_i)$

f) $E(\sum X_i) = \sum E(X_i)$

g) $Let X_1, X_2, ..., X_n$ be random variables which have mean μ. The sample mean is $\overline{X} = (\frac{\sum_{i=1}^{n} X_i}{n})$

h) $Var(aX + bY) = a^2 Var(X) + bVar(Y) + 2abCov(X, Y)$

i) $Var(X + Y) = Var(X - Y) = Var(X) + Var(Y)$ if X and Y are independent.

j) $\sum a_i^2 \sigma_1^2 + 2 \sum_{i<j} a_i a_j Cov(X_i, X_j)$

k) $Var(\sum a_i X_i) = \sum a_i^2 Var(X_i) if X_1, ..., X_n$ are independent

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

5 Continuity Correction

Only apply continuity correction when approximating from discrete to continious

- If $P(x = n)$ use $P(n - 0.5 < X < n + 0.5)$ (eg. $P(x = 5)$ want values around 5)
- If $P(x > n)$ use $P(X > n + 0.5)$ (eg. $P(x > 5)$ we don't want the 5)
- If $P(x \leq n)$ use $P(X < n + 0.5)$ (eg. $P(x \leq 5)$ we do want the 5)
- If $P(x < n)$ use $P(X < n - 0.5)$
- If $P(x \geq n)$ use $P(X > n - 0.5)$

6 Confidence Intervals

6.1 Two-sided Intervals

In general use $P(Z < c) = \frac{(1-100p\%)}{2} + 100p\%$

- For 99% confidence use $z = 2.58$
- For 98% confidence use $z = 2.326$
- For 95% confidence use $z = 1.96$
- For 90% confidence use $z = 1.645$

6.2 One-sided Intervals

- For 95% confidence use $[-\infty, 1.645]$

6.3 General

- $\hat{\mu} \pm c * \frac{\sigma}{\sqrt{n}}$

7 Z, t, Chi Calculations

Let n represent some number

- If $P(Z < -n)$ use $1 - P(Z < n)$ (so that we can use table)
- If $P(Z > -n)$ use $P(Z < n)$
- If $P(Z > n)$ use $1 - P(Z < n)$
- If $P(-n < Z < n) = 2P(Z < n) - 1$ (if you have to rearrange then add 1 first)