Statistics Cheat Sheet

Ch 1: Overview & Descriptive Stats

Populations, Samples and Processes

`Population:` well-defined collection of objects
`Sample:` a subset of the population
`Descriptive Stats:` summarize & describe features of data
`Inferential Stats:` generalizing from sample to population
`Probability:` bridge btwn descriptive & inferential techniques.
In probability, properties of the population are assumed known & questions regarding a sample taken from the population are posed and answered.
`Discrete and Continuous Variables:` A numerical variable is *discrete* if its set of possible values is at most countable.
A numerical value is *continuous* if its set of possible values is an uncountable set.
Probability: pop $\rightarrow$ sample
Stats: sample $\rightarrow$ pop

Measures of Location

For observations $x_1, x_2, \cdots, x_n$

`Sample Mean`	$\bar{x} = \frac{\sum_{1=1}^{n} x_i}{n}$
`Sample Median`	$\tilde{x} = (\frac{n+1}{2})^{\text{nth}}$ observation
`Trimmed Mean`	btwn $\tilde{x}$ and $\bar{x}$, compute by removing smallest and largest observations

Measures of Variability

`Range`	=lgst-smllst observation
`Sample Variance`, σ^2	$= \frac{\sum(x_i - \bar{x}))^2}{n-1} = \frac{S_{xx}}{n-1}$
S_{xx}	$= \sum x_i^2 - \frac{(\sum x_i)^2}{n}$
`Sample Standard Deviation`, σ	$= \sqrt{\sigma^2}$

Box Plots

Order the n observations from small to large. Separate the smallest half from the largest (If n is odd then $\tilde{x}$ is in both halves). The lower fourth is the median of the smallest half (upper fourth..largest..). A measure of the spread that is resistant to outliers is the *fourth spread* f_s given by $f_s =$ upper fourth- lower fourth. Box from lower to upper fourth with line at median. Whiskers from smallest to largest x_i.

Ch 2: Probability

Sample Space and Events

`Experiment`	activity with uncertain outcome
`Sample Space`($\mathcal{S}$)	the set of all possible outcomes
`Event`	any collection of outcomes in $\mathcal{S}$

Axioms, Interpretations and Properties of Probability

Given an experiment and a sample space $\mathcal{S}$, the objective probability is to assign to each event A a number $P(A)$, called the probability of event A, which will give a precise measure of the chance that A will occur. Behaves very much like norm.

Axioms & Properties of Probability:

1. $\forall A \in \mathcal{S}, 0 \leq P(A) \leq 1$
2. $P(\mathcal{S}) = 1$
3. If $A_1, A_2, \ldots$ is an infinite collection of disjoint events, $P(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i)$
4. $P(\emptyset) = 0$
5. $\forall A, P(A) + P(A') = 1$ from which $P(A) = 1 - P(A')$
6. For any two events $A, B \in \mathcal{S}$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
7. For any three events $A, B, C \in \mathcal{S}, P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Equally Likely Outcomes : $P(A) = \frac{N(A)}{N}$

Counting Techniques

`Product Rule for Ordered k-Tuples:` If the first element can be selected in n_1 ways, the second in n_2 ways and so on, then there are $n_1 n_2 \cdots n_k$ possible k-tuples.
`Permutations:` An ordered subset. The number of permutations of size k that can be formed from a set of n elements is $P_{k,n}$

$$P_{k,n} = (n)(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

`Combinations:` An unordered subset.

$$\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$

Conditional Probability

$P(A|B)$ is the conditional probability of A given that the event B has occurred. B is the conditioning event.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Multiplication Rule: $P(A \cap B) = P(A|B) \cdot P(B)$

Baye's Theorem

Let $A_1, A_2, \ldots, A_k$ be disjoint and exhaustive events (that partition the sample space). Then for any other event B
$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)$
$= \sum_{i=1}^{k} P(B|A_i)P(A_i)$

Independence

Two events A and B are `independent` if $P(A|B) = P(A)$ and are `dependent` otherwise.
A and B are `independent` iff $P(A \cap B) = P(A) \cdot P(B)$ and this can be generalized to the case of n mutually independent events.

Random Variables

`Random Variable`:	any function $X : \Omega \rightarrow \mathbb{R}$
`Prob Dist.`:	describes how the probability of Ω is distributed along the range of X
`Discrete rv`:	rv whose domain is at most countable
`Continuous rv`:	rv whose domain is uncountable and where $\forall c \in \mathbb{R}, P(X = c) = 0$
`Bernoulli rv`:	discrete rv whose range is $\{0, 1\}$

The *probability distribution of* X says how the total probability of 1 is distributed among the various possible X values.

1. Distributions

Discrete RVs

Probabilities assigned to various outcomes in $\mathcal{S}$ in turn determine probabilities associated with the values of any particular rv X.
`Probability Mass Fxn/Probability Distribution,(pmf):`

$$p(x) = P(X = x) = P(\forall w \in \mathcal{W} : X(w) = x)$$

Gives the probability of observing $w \in \mathcal{W} : X(w) = x$
The conditions $p(x) \geq 0$ and $\sum_{\text{all possible x}} p(x) = 1$ are required for any pmf.
`parameter:` Suppose p(x) depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a parameter of distribution. The collection of all probability distributions for different values of the parameter is called a family of probability distributions.

Cumulative Distribution Function

(To compute the probability that the observed value of X will be at most some given x)
`Cumulative Distribution Function(cdf):` $F(x)$ of a discrete rv variable X with pmf $p(x)$ is defined for every number x by

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y)$$

For any number $x, F(x)$ is the probability that the observed value of X will be at most x.
For discrete rv, the graph of $F(x)$ will be a step function- jump at every possible value of X and flat btwn possible values.
For any two number a and b with $a \leq b$:
$P(a \leq X \leq b) = F(b) - F(a^-)$
$P(a < X \leq b) = F(b) - F(a)$
$P(a \leq X \leq a) = F(a) - F(a^-) = p(a)$
$P(a < X < b) = F(b^-) - F(a)$
(where a^- is the largest possible X value strictly less than a)
Taking $a = b$ yields $P(X = a) = F(a) - F(a-1)$ as desired.
`Expected value or Mean Value`

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Describes where the probability distribution is centered and is just a weighted average of the possible values of X given their distribution. However, the sample average of a sequence of X values may not settle down to some finite number (harmonic series) but will tend to grow without bound. Then the distribution is said to have a *heavy tail*. Can make if difficult to make inferences about μ.
`The Expected Value of a Function`: Sometimes interest will focus on the expected value of some function $h(x)$ rather than on just $E(x)$.
If the RV X has a set of possible values D and pmf $p(x)$, then the expected value of any function $h(x)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$ is computed by

$$E[h(X)] = \sum_{D} h(x) \cdot p(x)$$

`Properties of Expected Value`:

$$E(aX+b) = a \cdot E(X) + b$$

`Variance of X`: Let X have pmf $p(x)$ and expected value μ. Then the $V(X)$ or σ_X^2 is

$$V(X) = \sum_D (x-\mu)^2 \cdot p(x) = E[(X-\mu)^2]$$

The standard deviation (SD) of X is $\sigma = \sqrt{\sigma}$
Alternatively,

$$V(X) = \sigma^2 = \big[\sum_D x^2 \cdot p(x)\big] - \mu^2 = E(X^2) - [E(X)]^2$$

`Properties of Variance`

1. $V(aX+b) = a^2 \cdot \sigma^2$
2. In particular, $\sigma_{aX} = |a| \cdot \sigma_x$
3. $\sigma_{X+b} = \sigma_X$

Continuous RVs

Probabilities assigned to various outcomes in $\mathcal{S}$ in turn determine probabilities associated with the values of any particular rv X. Recall: an rv X is continuous if its set of possible values is uncountable and if $P(X=c)=0 \quad \forall c.//$

`Probability Density Fxn/Probability Distribution,(pdf)`: $\forall a, b \in \mathbb{R}, a \leq b$

$$P(\forall w \in \mathcal{W} : a \leq X(w) \leq b) = \int_a^b f(x)dx$$

Gives the probability that X takes values between a and b. The conditions $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x) = 1$ are required for any pdf.

`Cumulative Distribution Function(cdf)`:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

For any number $x, F(x)$ is the probability that the observed value of X will be at most x.
By the continuity arguments for continuous RVs we have that

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b)$$

Other probabilities can be computed from the cdf $F(x)$:

$$P(X > a) = 1 - F(a)$$

$$P(a \leq X \leq b) = F(b) - F(a)$$

Furthermore, if X is a cont rv with pdf $f(x)$ and cdf $F(x)$, then at every x at which $F'(x)$ exists, $F'(x) = f(x)$.
`Median`($\tilde{\mu}$): is the 50th percentile st $F(\tilde{\mu}) = .5$. That is half the area under the density curve. For a symmetric curve, this is the point of symmetry.
`Expected/Mean Value`(μ or $E(X)$): of cont rv with pdf $f(x)$

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

If X is a cont rv with pdf $f(x)$ and $h(X)$ is any function of X then

$$E[h(X)] = \mu = \int_{-\infty}^{\infty} h(x) \cdot f(x)dx$$

`Variance`: of a cont rv X with pdf $f(x)$ and mean value μ is

$$\sigma_x^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx = E[(X-\mu)^2]$$

Alternatively,

$$V(X) = E(X^2) - [E(X)]^2$$

Discrete Distributions

The Binomial Probability Distribution

1) The experiment consists of n *trials* where n is fixed
2) Each trial can result in either success (S) or failure (F)
3) The trials are independent
4) The probability of success $P(S)$ is constant for all trials

Note that in general if the sampling is without replacement, the experiment will not yield independent trials. However, if the sample size (number of trials) n is at most 5% of the population, then the experiment can be analyzed as though it were exactly a binomial experiment.
`Binomial rv X`: = no of S's among the n trials
`pmf of a Binomial RV`:,

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad : x = 0, 1, 2, \ldots$$

`cdf for Binomal RV`: Values in Tble A.1

$$B(x; n, p) = P(X \leq x) = \sum_{y=0}^{x} b(y; n, p)$$

`Mean & Variance of X` If $X \sim Bin(n, p)$ then

$$E(X) = np \quad V(X) = npq$$

Negative Binomial Distribution

1) The experiment consists of independent trials
2) Each trial can result in either Success(S) or Failure(F)
3) The probability of success is constant from trial to trial
4) The experiment continues until a total of r successes have been observed, where r is a specified integer.

`RV Y`: = the no of trials before the rth success.
`Negative Binomial rv`: $X = Y - r$ the number of failures that precede the rth success. In contrast to the binomial rv, the number of successes is fixed while the number of trials is random.
`pmf of the negative binomial rv` : with parameters r = number of S's and $p = P(S)$ is

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x \qquad x = 0, 1, 2, \ldots$$

`Mean & Variance of negative binomial rv X`: with pmf $nb(x; r, p)$

$$E(X) = \frac{r(1-p)}{p} \quad V(X) = \frac{r(1-p)}{p^2}$$

Geometric Distribution

`RV X`: = the no of trials before the 1st success.
`pmf of the geometric rv` :

$$p(x) = q^{x-1}p$$

$$E(X) = \sum x q^{x-1} p = 1/p$$

The Poisson Probability Distribution

Useful for modeling rare events
1) independent: no of events in an interval is independent of no of events in another interval
2) Rare: no 2 events at once
3) Constant Rate: average events/unit time is constant ($\mu > 0$)
`RV X=` no of occurrence in unit time interval
`Possion distribution/ Poisson pmf`: of a random variable X with parameter $\mu > 0$ where

$$p(x; \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!} \quad x = 0, 1, 2, \ldots$$

`Binomial Approximation`: Suppose that in the binomial pmf $b(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. Then $b(x; n, p) \to p(x; \mu)$.
That is to say that in any binomial experiment in which n(the number of trials) is large and p(the probability of success) is small, then $b(x; n, p) \approx p(x; \mu)$, where $\mu = np$.
`Mean and Variance of X`: If X has probability distribution with parameter μ, then $E(X) = V(X) = \mu$

Continuous Distributions

The Normal Distribution, $X \sim N(\mu, \sigma^2)$

`PDF`: with parameters μ and σ where $-\infty < \mu < \infty$ and $0 < \sigma$

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \quad -\infty < x < \infty$$

We can then easily show that $E(X) = \mu$ and $V(X) = \sigma^2$.
`Standard Normal Distribution`: The specific case where $\mu = 0$ and $\sigma = 1$. Then

$$\text{pdf}: \quad \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{cdf}: \quad \Phi(z) = \int_{-\infty}^{z} \phi(u)du$$

`Standardization`: Suppose that $X \sim N(\mu, \sigma^2)$. Then

$$Z = (X - \mu)/\sigma$$

transforms X into standard units. Indeed $Z \sim N(0, 1)$.

$$P(a \leq X \leq b) = P\Big(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\Big) = \Phi\Big(\frac{b-\mu}{\sigma}\Big) - \Phi\Big(\frac{a-\mu}{\sigma}\Big)$$

`Independence`: If $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ and X and Y are independent, then $X \pm Y \sim N(\mu_x \pm \mu_y, \sigma_x^2 + \sigma_y^2)$
NOTE: By symmetry of the standard normal distribution, it follows that $\Phi(-z) = 1 - \Phi(z) \quad \forall z \in \mathbb{R}$