STAT230 - Study Notes

An open source note project

Original Author:
Daniel Chevalier (AKA: TopGunCoder)

This is an open source study note that was made to help with the understanding of STAT230 content.

To access the repository for this study note:

https://github.com/TopGunCoder/STAT230-StudyAide

The more people who contribute and pass this on, the more that yourself and future STAT230 students can benefit from this.

Pass it forward! :)

NOTE: These notes are taken directly from:
STAT 220/230 NOTES (2011-12 Edition)
originally by Chris Springer
edited by Jerry Lawless and Don Mcleish

Key Terms

8.2: Joint Probability Function:

$f(x_1, ..., x_k) = \frac{n!}{x_1!...x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$

8.3: The distribution of X_n

Suppose that the chain is started by randomly choosing a state for X_0 with distrubution $P[X_0 = i] = q_i, i = 1, 2, ..., N$. Then the distrubution of X_1 is guben by:
$P(X_1 = j) = \sum_{i=1}^{N} P(X_1 = j, X_0 = i)$
$= \sum_{i=1}^{N} P(X_1 = j|X_0 = i)P(X_0 = i)$
$\sum_{i=1}^{N} Pijqiy$

7.2: Linearity Properties of Expectation

1. For constants a and b
   $E[ag(x) + b] = aE[g(x)] + b$
2. Similarly for constants a and b and two functions g_1 and g_2, it is also easy to show:
   $E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$

7.4: Properties of Mean and Variance

If a and b are constants and $Y = aX + b$, then
$\mu_Y = a\mu_X + b$ and $\sigma_Y^2 = a^2\sigma_X^2$
(where μ_X and σ_X^2 are the mean and variabnce of X and μ_Y and σ_Y^2 are the mean and variance of Y).

8.4: Property of Multivariate Expectation

It is easily proved (make sure you can do this) that
$E[ag_1(X, Y) + bg_2(X, Y)] = aE[g_1(X, Y)] + bE[g_2(X, Y)]$
This can be extended beyond 2 functions g_1 and g_2, and beyond 2 variables X and Y

8.5: Results for Means

1. $E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y$, when a and b are constants. (This follows from the definition of expected value .) In particular, $E(X + Y) = \mu_X + \mu_Y$ and $E(X - Y) = \mu_X - \mu_Y$

2. Let a_i be constants (real numbers) and $E(x_i) = \mu_i$. Then $E(\sum a_i X_i) = \sum a_i \mu_i$. In particular, $E(\sum X_i) = \sum E(X_i)$.

3. Let $X_1, X_2, ..., X_n$ be randome variables which have mean μ. (You can imagine these being some sample results from and experiment such as recording the number of occupants in cars travelling over a toll bridge.) The sample mean is $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. Then $E(\overline{X}) = \mu$

8.5: Results for Covariance

1. $Cov(X, X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu)^2] = Var(X)$

2. $Cov(aX+bY, cU+dV) = acCov(X,U)+adCov(X,V)+bcCov(Y,U)+bdCov(Y,V)$ where a, b, c, and d are constants.

8.5: Results for Variance

1. **Variance of linear combination:**
$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$

2. **Variance of a sum of independent random variables**
Let X and Y be independent. Since $Cov(X,Y) = 0$, result 1. gives
$Var(X + Y) = \sigma_X^2 + \sigma_Y^2$;
i.e., for independent variables, the variance of a difference is the sum of the variances.

3. **Variance of a general linear combination:**
Let a_i be constants and $Var(X_i) = \sigma_i^2$. Then
$Var(\sum a_i X_i) = \sum a_i^2 \sigma_i^2 + 2\sum_{i<j} a_i a_j Cov(X_i, X_j)$.
This is a generalization of result 1. and can be proved using either of the methods used for 1.

4. **Variance of a linear combination of independent**
Special cases of result 3. are:

   a) If $X_1, X_2, ..., X_n$ are independent then $Cov(X_i, X_j) = 0$, so that
   $Var(\sum a_i X_i) = \sum a_i^2 \sigma_i^2$.
   b) If $X_1, X_2, ..., X_n$ are independent and all have the same variance σ^2, then
   $Var(\overline{X}) = \frac{\sigma^2}{n}$

8.5: Indicator Variables

The results for linear combinations of random variables provide a way of breaking up more complicated problems, involving mean and cariance, into simpler pieces using indicator variables; an indicator variable is just a binary variable (o or 1) that indicates whether or not some event occurs.

9.1: Continuous random variables

Continuous random variables have a range (set of possible values) and interval (or a collection of intervals) on the real number line. They have to be treated a little differently than discrete random variables because $P(X = x)$ is zero for each x.

9.1: Cumulative Distribution Function

For discrete random variables we defined the c.d.f, $F(x) = P(X \leq x)$ for continuous random variables as well as for discrete.

9.1: Properties of a probability density function

1. $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)\ \mathrm{d}x$. (This follwos from the definition of $f(x)$)
2. $f(x) \geq 0$. (Since $F(x)$ is non-decreasing, its derivatice is non-negative)
3. $\int_{-\infty}^{\infty} f(x)\ \mathrm{d}x = \int_{all\ x} f(x)\ \mathrm{d}x = 1$. (this is because $P(-\infty \leq X \leq \infty) = 1$)
4. $F(x) = \int_{-\infty}^{\infty} f(u)\ \mathrm{d}u$. (this is just property 1 with $a = -\infty$)

9.1: Defined Variables or Change of Variable

When we know the p.d.f or c.d.f. for a continuous random variable X we sometimes want to find the p.d.f. or c.d.f for some other random variable Y which is a function of X. The procedure for doing this is summarized below. It is based on the fact that the c.d.f $F_Y(y)$ for Y equals $P(Y \leq y)$, and this can be rewritten in terms of X since Y is a function of X. Thus:

1. Write the c.d.f. of Y as a function of X.
2. Use $F_X(x)$ to find $F_Y(y)$. Then if you want the p.d.f. $f_Y(y)$, you can differentiate the expression for $F_Y(y)$.

3. Find the range of values of y.

9.2: The probability density function and the cumulative distribution function

Since all points are equally likely (more precisely, intervals contained in $[a, b]$ of a given length, say 0.01, all have the same probability), the probability density function must be a constant $f(x) = l; a \leq x \leq b$ fro some constant k. To make $\int_a^b f(x)\,\mathrm{d}x = 1$, we require $k = \frac{1}{b-a}$.
Therefore $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$

$$F(x) = \begin{cases} 0 & : for\ x < a \\ \int_a^x \frac{1}{b-a}\,\mathrm{d}x & : for\ a \leq x \leq b \\ 1 & : for\ x > b \end{cases}$$

9.2: Mean and Variance

$\mu = \frac{b+a}{2}$
$E(X^2) = \frac{b^2+ab+a^2}{3}$
$\sigma^2 = \frac{(b-a)^2}{12}$

9.3: Exponential Distribution

The continuous random variable X is siad to have an **exponential distribution** if its p.d.f. is of the form
$f(x) = \lambda e^{-\lambda x} \qquad x > 0$

9.3: p.d.f and c.d.f of Exponential Distribution

$F(x) = 1 - \frac{\mu^0 e^{-\mu}}{0!} = 1 - e^{-\mu}$.
$f(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(x) = \lambda e^{-\lambda x}$; for $x > 0$

9.3: The Memoryless Property of the Exponential Distribution

$P(X > a + b | X > b) = P(X > a)$

9.5: Normal Distribution p.d.f

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \qquad \infty < x < \infty$
$E(x) = \mu$

$Var(X) = \sigma^2$
So it's p.d.f is written as
$X \sim N(\mu, \sigma^2)$

9.5: The Cumulative Distribution Function of the Normal Distribution

The c.d.f. of the normal distribution $N(\mu, \sigma^2)$ is
$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2}\ \mathrm{d}y$

9.5: Mean,Variance, and Moment Generating Function of a normal distrubution

$E(X) = \mu$
$Var(X) = \sigma^2$
$M_X(t) = E(e^{Xt}) = e^{\mu t + \sigma^2 t^2/2}$

9.5: Gaussian Distribution

The normal distribution is also known as the Gaussian distribution. The notation $X \sim G(\mu, \sigma)$ means that X has Gaussian (normal) distribution with meanμ and standard deviation σ. So, for example, if $X \sim N(1, 4)$ then we could also write $X \sim G(1, 2)$.

9.5 Linear Combinations of Independent Normal Random Variables

1. Let $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, where a and b are constant real numbers. Then $Y \sim N(a\mu + b, a^2\sigma^2)$

2. Let $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ be independent, and let a and b be constants.
   Then $aX + bY \sim N(a\mu + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$.
   In general if $X_i \sim N(\mu_i, \sigma_i^2)$ are independent and a_i are constants, then $\sum a_i X_i \sim N(\sum a_i\mu_i, \sum a_i^2\sigma_i^2)$.

3. let $X_1, ..., X_n$ be independent $N(\mu, \sigma^2)$ random variables.
   Then $\sum X_i \sim N(n\mu, n\sigma^2)$ and $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Theorems

7.2: Theorem 16

Suppose the random variable X has probability function $f(x)$. Then the **expected vaue** of some function $g(x)$ of X is given by
$E[g(x)] = \sum_{allx} g(x)f(x)$

7.5: Theorem 20

Let the random variable X gave m.g.f. $M(t)$. Then
$E(X^r) = M^{(r)}(0)\ r = 1, 2, ...$
where $M^{(r)}(0)$ stands for $d^r M(t)/dt^r$ evaluated at t=0

8.4: Theorem 27

If X and Y are independent then
$Cov(X, Y) = 0$

8.4: Theorem 28

Suppose random variables and Y are independent. Then, if $g_1(X)$ and $g_2(Y)$ are any two functions,
$E[g_1(X)g_2(Y)] = E[g_1(X)]E(g_2(Y)]$

8.6: Theorem 31

The **moment generating function of the sum of independent random variables** is the product of the individual moment generating functions.

9.4: Theorem 35

If F is an arbitrary c.d.f. and U is uniform on $[0, 1]$ then the random variable defined by $X = F^{-1}(U)$ has c.d.f. $F(x)$.

9.5: Theorem 36

Let $X \sim N(\mu, \sigma^2)$ and define $Z = \frac{(X-\mu)}{\sigma}$. Then $Z \sim N(0, 1)$ and
$F_X(x) = P(X \leq x) = F_Z(\frac{x-\mu}{\sigma})$.

9.6: Theorem 37

If $X_1, X_2, ..., X_n$ are independent random variables all having the same distribtion, with mean μ and variance σ^2, then as $n \to \infty$, the c.d.f. of the the random variable

$\frac{\sum X_i - n\mu}{\sigma\sqrt{n}}$

approaches the $N(0,1)$ c.d.f. Similarly, the c.d.f. of

$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

approaches the standard normal c.d.f.

9.6: Theorem 38

Let X have a binomial distribution, $Bi(n,p)$. Then for n large, the random variable

$W = \frac{X-np}{\sqrt{np(1-p)}}$

is approximately $N(0,1)$.

Definitions

7.1: Definition 13

The **median** of a sample is a value such that half the results are below it and half above it, when the results are arranged in numerical order.

7.1: Definition 14

The **mode** of the sample is the value which occurs most often. There is no guarantee there will be only a single mode.

7.2: Definition 15

The **expected value** (also called the mean or the expectation of a discrete random variable X with probability function $f(x)$ is

$E(X) = \sum_{allx} xf(x)$

7.4: Definition 17

The **variance** of a r.v X is $E[(X-\mu)^2]$, and is denoted by σ^2 or by $Var(X)$

7.4: Definition 18

The **standard deviation** of a random variable X is $\sigma = \sqrt{E[(x-\mu)^2]}$

1. $\sigma^2 = E(X^2) - \mu^2$
2. $\sigma^2 = E[X(X-1)] + \mu - \mu^2$

7.5: Definition 19

Consider a discrete random variable X with probability function $f(x)$. The **moment generating function (m.g.f)** of X is defined as
$M(t) = E(e^{tX}) = \sum_x e^{tx} f(x)$
We will assume that the moment generating function is defined and finite for values of t in an interval around 0 (i.e. for some $a > 0$, $\sum_x d^{tX} f(x) < \infty$ for all $t \in [-a, a]$).

8.1: Definition 21

X and Y are **independent** random variables iff $f(x, y) = f_1(x) f_2(y)$ for all values (x, y)

8.1: Definiton 22

In general, $X_1, ..., X_n$ are independent random variables iff
$f(x_1, ..., x_n) = f_1(x_1) f_2(x_2) ... f_n(x_n)$ for all $x_1, ..., x_n$

8.1: Definition 23

The conditinal probability function of X given $Y = y$ is $f(x|y) = \frac{f(x,y)}{f_2(y)}$.
Similarly, $f(y|x) = \frac{f(x,y)}{f_1(x)}$ (provided, of course, the denominator is not zero).

8.3: Definition 24

A stationary distribution of a Markov chain is the column vector (π say) of probabilities of the individual states such that $\pi^T P = \pi^T$.

8.4: Definition 25

$E[g(X, Y)] = \sum_{all(x,y)} g(x, y) f(x, y)$
and
$E[g(X_1, ..., X_n)] = \sum_{all(x_1,...,x_n)} g(x_1, ..., x_n) f(x_1, ..., x_n)$

8.4: Definition 26 - Covariance

The **Covariance** of X and Y, denoted Cov(X,Y) or σ_{XY}, is
$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$
For Calculation purposes this definition is usually harder to use than the formula which follows:
$Cov(X, Y) = E(XY) - E(X)E(Y)$
(Note that this is the result of using **"8.4: Property of Multivariate Expectation"**)

8.4: Definition 29 - Correlation Coefficient

The **correlation coefficient** of X and Y is:
$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

8.6: Definition 30

The **joint moment generating function** of (X, Y) is
$M(s, t) = E\{e^{sX+tY}\}$
Recall that if X, Y happen to be independent, $g_1(X)$ and $g_2(Y)$ are any two functions,
$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$.
and so with $g_1(X) = e^{sX}$ and $g_2(Y) = e^{tY}$ we obtain, for independent random variables X, Y
$M(s, t) = M_X(s)M_Y(t)$

9.1: Definition 32 - Probability Density Function

Th **probability density functions** (p.d.f.) $f(x)$ for a continuous random variable X is the derivative
$f(x) = \frac{dF(x)}{dx}$
where $F(x)$ is the c.d.f for X.

9.1: Definition 33 - Extension of Expectation, Mean, and Variance to Continuous Distributions

When X is continuous, we still define
$E(g(X)) = \int_{all\ x} g(x)f(x)\,\mathrm{d}x$.
With this definition, all of the earlier properties of expected value and variance still hold; for example with $\mu = E(X)$,

$$\sigma^2 = Var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2.$$

9.3: Definition 34 - The Gamma Function

$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} \mathrm{d}x$ is called the gamma function of a, wherer $a > 0$
Properties of the gamma function:

1. $\Gamma(a) = (a-1)\Gamma(a-1)$ for $a > 1$

2. $\Gamma(a) = (a-1)!$ if a is a positive integer

3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$