Midterm Review

Sta 111

Colin Rundel

June 3, 2014

1 Midterm Review

- Probability
- Distributions
- Functions of Random Variables
- Moments of Distributions
- LLN and CLT
- Continuous Random Variables
- Order Statistics
- Continuous Distributions
- Joint Distributions of Discrete RVs
- Joint Distributions of Continuous RVs
- Bivariate Normal

Rules of Probability

(1) Non-negative:

$$P(E) >= 0$$

(2) Addition:

$$P(E \cup F) = P(E) + P(F) \text{ if } EF = \emptyset$$

(2)' Countable Addition:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) \text{ if } E_i E_j = \emptyset \text{ for } i \neq j$$

(3) Total one:

$$P(\Omega) = 1$$

Useful Identities

Commutativity & Associativity:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$
$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$
$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

*Think of union as addition and intersection as multiplication: $(A + B)C = AC + BC$

DeMorgan's Rules:

$$\text{not } (A \text{ and } B) = (\text{not } A) \text{ or } (\text{not } B)$$
$$\text{not } (A \text{ or } B) = (\text{not } A) \text{ and } (\text{not } B)$$

Useful Identities, cont.

Complement Rule:

$$P(\text{not } A) = P(A^c) = 1 - P(A)$$

Difference Rule:

$$P(B \text{ and not } A) = P(BA^c) = P(B) - P(A) \text{ if } A \subseteq B$$

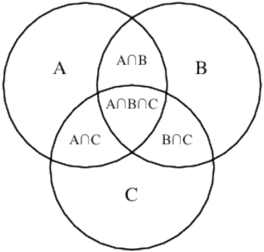
Inclusion-Exclusion:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Generalized Inclusion-Exclusion

For the case of $n = 3$:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A\ B) - P(A\ C) - P(B\ C) + P(A\ B\ C)$$



$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i \le n} P(E_i) - \sum_{i<j\le n} P(E_i E_j) + \sum_{i<j<k\le n} P(E_i E_j E_k) - \ldots + (-1)^{n+1} P(E_1 \ldots E_n)$$

Equally Likely Outcomes

$$P(E) = \frac{\#(E)}{\#(\Omega)} = \sum_i \frac{1_{\omega_i \in E}}{\#(\Omega)}$$

Sampling:

- Sampling with replacement
- Sampling without replacement
- Pólya urn model

Conditional Probability

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

Multiplication rule:

$$P(A \cap B) = P(A|B)P(B)$$

Total probability:
For a partition $B_1, \ldots, B_n$ of Ω,

$$P(A) = P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)$$

Independence

We defined events A and B to be independent when

$$P(A \cap B) = P(A)P(B)$$

which also implies that

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

Not to be confused with mutually exclusive events where

$$P(A \cap B) = 0$$

Bayes' Rule

For a partition $B_1, \ldots, B_n$ of all possible outcomes,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}$$

Generalizing Conditional Probability

For three events:

$$P(A \cap B \cap C) = P(A \cap B)P(C|A,B) = P(A)P(B|A)P(C|A,B)$$

For n events:

$$P(\cap A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1,A_2)\cdots P(A_n|A_1,\ldots,A_{n-1})$$

Combinations & Permutations

Selecting k items from a collection of n then,

If we don't care about order - Combinations (Binomial Coefficient):

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

$$\sum_{0}^{n} \binom{n}{k} = 2^n$$

If we do care about order - Permutations:

$$\frac{n!}{(n-k)!}$$

Bernoulli Distribution

Let X be a random variable that that takes the value 1 upon success or 0 upon failure of a single trial where the probability of success is given by p, $X \sim \text{Bern}(p)$

$$P(X = k|p) = f(k|p) = \begin{cases} p & \text{if } k = 1, \\ 1 - p & \text{if } k = 0. \end{cases}$$

$$E(X) = p$$
$$Var(X) = p(1 - p)$$
$$\text{Mode}(X) = \begin{cases} 0 & \text{if } q > p \\ 0, 1 & \text{if } q = p \\ 1 & \text{if } q < p \end{cases}$$

Binomial Distribution

Let X be a random variable that reflects the *number of successes* in a *fixed number*, n of *independent trials* with the *same probability of success*, p, $X \sim \text{Binom}(n, p)$

$$P(X = k) = f(k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned}
\text{range}(X) &= \{0, 1, \ldots, n\} \\
E(X) &= np \\
Var(X) &= np(1-p) \\
\text{Mode}(X) &= \{\lceil np - q \rceil, \lfloor np + p \rfloor\}
\end{aligned}$$

Alternative view - $X = \sum_{i=1}^{n} Y_i$ where $Y_i \sim \text{Bern}(p)$

Poisson Distribution

Let X be a random variable reflecting the number of events in a given period where the expected number of events in that interval is λ then the probability of k occurrences $(k \geq 0)$ in the interval is given by the Poisson distribution, $X \sim \text{Pois}(\lambda)$

$$P(X = k) = f(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\begin{aligned} \text{range}(X) &= \{0, 1, \ldots, \infty\} \\ E(X) &= \lambda \\ Var(X) &= \lambda \\ \text{Mode}(X) &= \begin{cases} \lambda - 1, \lambda & \text{if } \lambda = \lceil \lambda \rceil \\ \lceil \lambda \rceil - 1 & \text{otherwise} \end{cases} \end{aligned}$$

Can be used to approximate Binomial distribution when p is very small

Hypergeometric Distribution

Let X be a random variable reflecting the number of successes in n draws without replacement from a finite population of size N with m desired items then the probability of k successes is given by the Hypergeometric distribution, $X \sim \text{Hypergeo}(N, m, n)$

$$P(X = k) = f(k|N, m, n) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

$$\begin{aligned}
\text{range}(X) &= \{\max(0, n + m - N), \ldots, \min(m, n)\} \\
E(X) &= \frac{mn}{N} \\
Var(X) &= n\frac{m}{N}\frac{N-m}{N}\frac{N-n}{N-1}
\end{aligned}$$

Geometric Distribution - Ver. 1

Let X be a random variable reflecting the number failures of independent Bernoulli trials, with probability of success p, needed before observing the first success. Then the probability of k failures before the first success is given by the Geometric distribution, $Y \sim \text{Geo}(p)$

$$P(X = k) = f(k|p) = p(1-p)^k$$

$$\text{range}(X) = \{0, 1, 2, \ldots, \infty\}$$
$$E(X) = \frac{1-p}{p}$$
$$Var(X) = \frac{1-p}{p^2}$$

Geometric Distribution - Ver. 2

Let X be a random variable reflecting the number of independent Bernoulli trials, with probability of success p, needed to observe the first success. Then the probability of the first success occuring on the x^{th} trial is given by the Geometric distribution, $Y \sim \text{Geo}(p)$

$$P(X = x) = f(x|p) = p(1-p)^{x-1}$$

$$\begin{aligned} \text{range}(X) &= \{1, 2, \ldots, \infty\} \\ E(X) &= \frac{1}{p} \\ Var(X) &= \frac{1-p}{p^2} \end{aligned}$$

Negative Binomial Distribution

Let X be a random variable reflecting the total number of successes before the rth failure where each trial is an independent Bernoulli trial with p probability of success. Then the probability of k successes is given by the Negative Binomial distribution, $X \sim \text{NB}(r, p)$

$$P(X = k) = f(k|r,p) = \binom{k+r-1}{k} p^k (1-p)^r$$

$$\begin{aligned} \text{range} &= \{0, 1, 2, \ldots, \infty\} \\ E(X) &= \frac{rp}{(1-p)} \\ Var(X) &= \frac{rp}{(1-p)^2} \end{aligned}$$

Alternative view - $X = Z_1 + Z_2 + \cdots + Z_r$ where $Z_1, \cdots, Z_r \sim Geo(p)$.

de Moivre-Laplace Limit Theorem

When n is large enough the Binomial distribution will have an approximately Normal Distribution.

- Approximation is usually considered reasonable when $np \geq 10$ and $nq \geq 10$
- de Moivre and Laplace where the first to identify this pattern and characterize the shape of the curve
- Special case of the Central Limit Theorem

Normal Distribution

If X is random variable with a normal distribution with a mean μ and variance σ^2, $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$P(X = x) = f(x) = \frac{1}{\sqrt{2\pi}\sigma} \, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

$$\begin{aligned} \text{range}(X) &\in (-\infty, \infty) \\ E(X) &= \mu \\ Var(X) &= \sigma^2 \\ \text{Mode}(X) &= \mu \end{aligned}$$

Unit Normal Distribution

The unit normal distribution is a special case of the normal distribution where $\mu = 0$ and $\sigma = 1$, $Z \sim \mathcal{N}(0, 1)$.

$$P(Z = z) = \phi(z) = \frac{1}{\sqrt{2\pi}}\; e^{-\frac{1}{2}z^2}$$

$$\begin{aligned} \text{range}(X) &\in (-\infty, \infty) \\ E(X) &= 0 \\ Var(X) &= 1 \\ \text{Mode}(X) &= 0 \end{aligned}$$

Properties of the Unit Normal Distribution

The area under the unit normal curve from $-\infty$ to a is given by

$$P(z \le a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2} dt = \Phi(a)$$

The area under the unit normal curve from a to ∞ is given by

$$P(z \ge a) = \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-t^2/2} dt = 1 - \Phi(a)$$

The area under the unit normal curve from a to b where $a \le b$ is given by

$$P(a \le z \le b) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-t^2/2} dt = \Phi(b) - \Phi(a)$$

Standardizing Normal Distributions

Let X be a normally distributed random variable with mean μ and variance σ^2 then we define the random variable Z such that

$$Z = \left(\frac{X-\mu}{\sigma}\right)$$

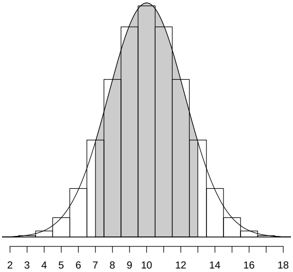
$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{E(X)-\mu}{\sigma} = 0$$

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{Var(X)}{\sigma^2} = 1$$

$$P(a \le x \le b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Improving the Discrete to Continuous Approximation

When approximating the Binomial distribution with the Normal distribution we miss a fraction of the probability



Continuity Correction gives better results:

$$P(a-1/2 \leq x \leq b+1/2) = \int_{a-1/2}^{b+1/2} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx = \Phi\left(\frac{b+1/2-\mu}{\sigma}\right) - \Phi\left(\frac{a-1/2-\mu}{\sigma}\right)$$

Properties of Expected Value

- $E(c) = c$
- $E(I_A) = P(A)$
- $E[g(X)] = \sum_{\text{all } x} g(x)\ P(X = x)$
- $E(cX) = cE(x)$
- $E(X + Y) = E(X) + E(Y)$
- $E(XY) = E(X)E(Y)$ if X and Y are independent.

Properties of Variance

- $Var(aX) = a^2 Var(X)$
- $Var(X + c) = Var(X)$
- $Var(c) = 0$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $Var(aX + bY + c) = a^2\ Var(X) + b^2\ Var(Y) + 2ab\ Cov(X, Y)$
- $Var\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n Cov(c_i X_i, c_j X_j)$
  $= \sum_{i=1}^n c_i^2 Var(X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n c_i c_j Cov(X_i, X_j)$

Moments

Raw moment:

$$\mu'_n = E(X^n)$$

Central moment:

$$\mu_n = E[(X - \mu)^2]$$

Normalized / Standardized moment:

$$\frac{\mu_n}{\sigma^n}$$

Moment Generating Function

The moment generating function of a random variable X is defined for all real values of t by

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} P(X = x) & \text{If } X \text{ is discrete} \\ \int_x e^{tx} P(X = x) dx & \text{If } X \text{ is continuous} \end{cases}$$

This is called the moment generating function because we can obtain the raw moments of X by successively differentiating $M_X(t)$ and evaluating at $t = 0$.

$$M_X(0) = E[e^0] = 1 = \mu_0'$$

$$M_X'(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt} e^{tX}\right] = E[Xe^{tX}]$$
$$M_X'(0) = E[Xe^0] = E[X] = \mu_1'$$

$$M_X''(t) = \frac{d}{dt} M_X'(t) = \frac{d}{dt} E[Xe^{tX}] = E\left[\frac{d}{dt}(Xe^{tX})\right] = E[X^2 e^{tX}]$$
$$M_X''(0) = E[X^2 e^0] = E[X^2] = \mu_2'$$

Moment Generating Function - Properties

If X and Y are independent random variables then the moment generating function for the distribution of $X+Y$ is

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$$

Similarly, the moment generating function for S_n, the sum of iid random variables $X_1, X_2, \ldots, X_n$ is

$$M_{S_n}(t) = [M_{X_i}(t)]^n$$

Markov's and Chebyshev's Inequalities

For any random variable $X \geq 0$ and constant $a > 0$ then

Markov's Inequality:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Chebyshev's Inequality:

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$$

LLN and CLT

Law of large numbers:

$$\lim_{n\to\infty} \frac{S_n - n\mu}{n} = \lim_{n\to\infty} (\bar{X}_n - \mu) \to 0$$

Central Limit Theorem:

$$\lim_{n\to\infty} \frac{S_n - n * mu}{\sqrt{n}} = \lim_{n\to\infty} \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

$$P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = P\left(a \le \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \le b\right) \approx \Phi(b) - \Phi(a)$$

CLT cont.

Law of large numbers:
Central Limit Theorem:
Let $X_1, X_2, \ldots, X_n \sim D$ where $E(D) = \mu$ and $Var(D) = \sigma^2$ then

$$S_n = X_1 + X_2 + \cdots + X_n \sim N(n\mu,\ n\sigma^2)$$

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim N(\mu,\ \sigma^2/n)$$

Cumulative Distribution Function

We have seen a variety of problems where we find $P(X <= x)$ or $P(X > x)$ etc. The former is given a special name - the cumulative distribution function.

If X is discrete with probability mass function $f(x)$ then

$$P(X \leq x) = F(x) = \sum_{z=-\infty}^{x} f(z)$$

If X is continuous with probability density function $f(x)$ then

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

CDF is defined for for all $-\infty < x < \infty$ and follows the following rules:

- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- $x < y \Rightarrow F(x) < F(y)$

Probability Density Function

For a continuous probability distribution

$$P(X = x) = 0 \text{ for all } x$$

As such we define the probability density function to be

$$f_X(x) = \lim_{\epsilon \to 0} P(X \in [x, x + \epsilon])/\epsilon$$

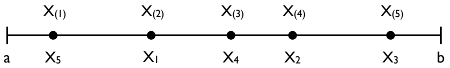
A pdf is defined for for all $-\infty < x < \infty$ and follows the following rules:

- $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- $\int_{-\infty}^{x} f_X(t)dt = F_X(x) \;\Leftrightarrow\; f_X(x) = \frac{d}{dx}F_X(x)$
- $f_X(x) \geq 0$ for all x

Order Statistics

Let X_1, X_2, X_3, X_4, X_5 be iid random variables with a distribution F with a range of (a, b). We can relabel these X's such that their labels correspond to arranging them in increasing order so that

$$X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq X_{(4)} \leq X_{(5)}$$



In the case where the distribution F is continuous we can make the stronger statement that

$$X_{(1)} < X_{(2)} < X_{(3)} < X_{(4)} < X_{(5)}$$

Since $P(X_i = X_j) = 0$ for all $i \neq j$ for continuous random variables.

Order Statistics, cont.

For $X_1, X_2, \ldots, X_n$ iid random variables X_k is the kth smallest X, usually called the kth order statistic.

$X_{(1)}$ is therefore the smallest X and

$$X_{(1)} = \min(X_1, \ldots, X_n)$$

Similarly, $X_{(n)}$ is the largest X and

$$X_{(n)} = \max(X_1, \ldots, X_n)$$

Distributions of order statistics

For $X_1, X_2, \ldots, X_n$ iid continuous random variables with pdf f and cdf F then

$$f_{(1)}(x) = nf(x)(1-F(x))^{n-1}$$
$$f_{(k)}(x) = nf(x)\binom{n-1}{k-1}F(x)^{k-1}(1-F(x))^{n-k}$$
$$f_{(n)}(x) = nf(x)F(x)^{n-1}$$

$$F_{(1)}(x) = 1-(1-F(x))^n$$
$$F_{(n)}(x) = F(x)^n$$

Uniform Distribution

If X is a random variable with constant density on (a, b) then X is said to be Uniformly distributed on (a, b), $X \sim \text{Unif}(a, b)$, then

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases} \qquad F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } x > b \end{cases}$$

$$E(X) = \frac{b+a}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

Normal Distribution

If X is random variable with a normal distribution with a mean μ and variance σ^2, $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \; e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$
$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$E(X) = \mu$$
$$Var(X) = \sigma^2$$
$$\text{Mode}(X) = \mu$$

Exponential Distribution

Let X be a random variable that reflects the time between events which occur continuously with a given rate λ, $X \sim \text{Exp}(\lambda)$

$$f(x|\lambda) = \lambda e^{-\lambda x}$$
$$P(X \le x) = F(x|\lambda) = 1 - e^{-\lambda x}$$

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} = \left(\frac{\lambda}{\lambda - t}\right)$$

$$E(X) = \lambda^{-1} \qquad E(X^n) = \frac{n!}{\lambda^n}$$
$$Var(X) = \lambda^{-2} \qquad Median(X) = \frac{\log 2}{\lambda}$$

Memoryless property - $P(X > s + t | X > s) = P(X > t)$

Minimum of Exponentials - $\min(X_1, \ldots, X_n) \sim \text{Exp}(\lambda_1 + \cdots + \lambda_n)$

Gamma Function

Based on the relationship for the nth raw moment, μ'_n, of an exponential distribution

$$E(X^n) = \frac{n!}{\lambda^n}$$

Let set $\lambda = 1$ and define an new value $\alpha = n + 1$

$$E(X^{\alpha-1}) = (\alpha - 1)!$$

$$\int_0^\infty x^{\alpha-1}e^{-x}dx = (\alpha - 1)!$$

$$\Gamma(\alpha) \equiv \int_0^\infty x^{\alpha-1}e^{-x}dx = (\alpha - 1)!$$

Using a tradition definition of the factorial it only makes sense when $n \in \mathbb{N}$ but we can use this new definition of the Gamma function $\Gamma(\alpha)$ for any $\alpha \in \mathbb{R}^+$.

Commonly used with the Gamma, Beta, and negative binomial distribution to generalize a parameter to $\mathbb{R}^+$.

Erlang Distribution

Let X reflect the time until the nth event occurs when the events occur according to a Poisson process with rate λ, $X \sim \text{Er}(n, \lambda)$

$$f(x|n,\lambda) = \frac{e^{-\lambda x}\lambda^n x^{n-1}}{(n-1)!}$$

$$F(x|n,\lambda) = \sum_{j=n}^{\infty} \frac{e^{-\lambda x}(\lambda x)^j}{j!}$$

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$$

$$E(X) = n/\lambda$$

$$Var(X) = n/\lambda^2$$

Gamma Distribution

We can generalize the Erlang distribution by using the gamma function instead of the factorial function, thereby allowing for $\mathbb{R}^+$ values of n. Often the distribution is reparameterized such that $\theta = 1/\lambda$, $X \sim$ Gamma(n, θ).

$$f(x|n,\theta) = \frac{e^{-x/\theta} x^{n-1}}{\theta^n \Gamma(n)}$$
$$F(x|n,\theta) = \frac{\int_0^x e^{-t/\theta} t^{n-1}\ dt}{\theta^n \Gamma(n)} = \frac{\gamma(n, x/\theta)}{\Gamma(n)}$$

$$M_X(t) = \left(\frac{1}{1-\theta t}\right)^n$$

$$E(X) = n\theta$$
$$Var(X) = n\theta^2$$

Beta Distribution

If $X \sim \text{Beta}(r, s)$ then

$$f(x) = \frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1}$$
$$F(x) = \int_0^x \frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1} dx = \frac{B_x(r,s)}{B(r,s)}$$

$$B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1} dx = \frac{(r-1)!(s-1)!}{(r+s-1)!} = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$
$$B_x(r,s) = \int_0^x x^{r-1}(1-x)^{s-1} dx$$

$$E(X) = \frac{r}{r+s}$$
$$Var(X) = \frac{rs}{(r+s)^2(r+s+1)}$$

Joint Distributions - Example

Draw two socks at random, without replacement, from a drawer full of twelve colored socks:

6 black, 4 white, 2 purple

Let B be the number of Black socks, W the number of White socks drawn, then the distributions of B and W are given by:

	0	1	2
P(B=k)	$\frac{6}{12}\frac{5}{11} = \frac{15}{66}$	$2\frac{6}{12}\frac{6}{11} = \frac{36}{66}$	$\frac{6}{12}\frac{5}{11} = \frac{15}{66}$
P(W=k)	$\frac{8}{12}\frac{7}{11} = \frac{28}{66}$	$2\frac{4}{12}\frac{8}{11} = \frac{32}{66}$	$\frac{4}{12}\frac{3}{11} = \frac{6}{66}$

Note - $B \sim \text{HyperGeo}(12, 6, 2) = \frac{\binom{6}{k}\binom{6}{2-k}}{\binom{12}{2}}$ and $W \sim \text{HyperGeo}(12, 4, 2) = \frac{\binom{4}{k}\binom{8}{2-k}}{\binom{12}{2}}$

Joint Distributions - Example, cont.

Let B be the number of Black socks, W the number of White socks drawn, then the distributions of B and W are given by:

		W: 0	W: 1	W: 2	
B	0	$\frac{1}{66}$	$\frac{8}{66}$	$\frac{6}{66}$	$\frac{15}{66}$
B	1	$\frac{12}{66}$	$\frac{24}{66}$	0	$\frac{36}{66}$
B	2	$\frac{15}{66}$	0	0	$\frac{15}{66}$
		$\frac{28}{66}$	$\frac{32}{66}$	$\frac{6}{66}$	$\frac{66}{66}$

$$P(B=b, W=w) = \frac{\binom{6}{b}\binom{4}{w}\binom{2}{2-b-w}}{\binom{12}{2}}$$

Marginal Distribution

Note that the column and row sums are the distributions of B and W respectively.

$$P(B = b) = P(B = b, W = 0) + P(B = b, W = 1) + P(B = b, W = 2)$$

$$P(W = w) = P(B = 0, W = w) + P(B = 1, W = w) + P(B = 2, W = w)$$

These are the *marginal* distributions of B and W. In general,

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_y P(X = x|Y = y)P(Y = y)$$

Conditional Distribution

Conditional distributions are defined as we have seen previously with

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{\text{joint pmf}}{\text{marginal pmf}}$$

Therefore the pmf for white socks given no black socks were drawn is

$$P(W = w|B = 0) = \frac{P(W = w, B = 0)}{P(B = 0)} = \begin{cases} \frac{1}{66} \Big/ \frac{15}{66} = \frac{1}{15} & \text{if } W = 0 \\ \frac{8}{66} \Big/ \frac{15}{66} = \frac{8}{15} & \text{if } W = 1 \\ \frac{6}{66} \Big/ \frac{15}{66} = \frac{6}{15} & \text{if } W = 2 \end{cases}$$

Expectation of Joint Distributions

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) P(X=x, Y=y)$$

For example we can define $g(x,y) = x \cdot y$ then

$$\begin{aligned} E(BW) =& (0 \cdot 0 \cdot 1/66) + (0 \cdot 1 \cdot 8/66) + (0 \cdot 2 \cdot 6/66) \\ &+ (1 \cdot 0 \cdot 12/66) + (1 \cdot 1 \cdot 24/66) + (1 \cdot 2 \cdot 0/66) \\ &+ (2 \cdot 0 \cdot 15/66) + (2 \cdot 1 \cdot 0/66) + (1 \cdot 2 \cdot 0/66) \\ =& 24/66 = 4/11 \end{aligned}$$

Note that $E(BW) \neq E(B)E(W)$ since

$$\begin{aligned} E(B)E(W) =& (0 \cdot 15/66 + 1 \cdot 36/66 + 2 \cdot 15/66) \\ &\times (0 \cdot 28/66 + 1 \cdot 32/66 + 2 \cdot 6/66) \\ =& 66/66 \times 44/66 = 2/3 \end{aligned}$$

This implies that B and W are not independent.

Expectation of Conditional Probability

Works like any other distribution

$$E(X|Y=y) = \sum_x xP(X=x|Y=y)$$

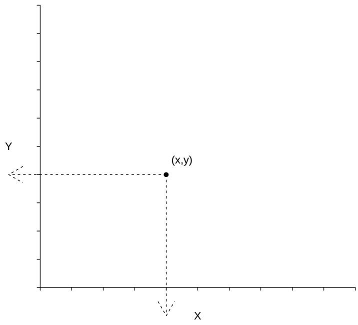
Therefore we can calculating things like conditional mean and variance,

$$E(W|B=0) = 0\cdot 1/15 + 1\cdot 8/15 + 2\cdot 6/15 = 20/15 = 1.333$$
$$E(W^2|B=0) = 0^2\cdot 1/15 + 1^2\cdot 8/15 + 2^2\cdot 6/15 = 32/15 = 2.1333$$

$$\begin{aligned} Var(W|B=0) &= E(W^2|B=0) - E(W|B=0)^2 \\ &= 32/15 - (4/3)^2 = 16/45 = 0.3556 \end{aligned}$$

Joint CDF

$$F(x,y) = P[X \le x, Y \le y]$$
$$= P[(X,Y) \text{ lies south-west of the point } (x,y)]$$



Joint CDF, cont.

The joint Cumulative distribution function follows the same rules as the univariate CDF,

Univariate definition:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(z)dz$$

$$\lim_{x \to -\infty} F(x) = 0 \qquad \lim_{x \to \infty} F(x) = 1 \qquad x \le y \Rightarrow F(x) \le F(y)$$

Bivariate definition:

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x,y)\; dx\; dy$$

$$\lim_{x,y \to -\infty} F(x,y) = 0 \qquad \lim_{x,y \to \infty} F(x,y) = 1 \qquad x \le x', y \le y' \Rightarrow F(x,y) \le F(x',y')$$

Marginal Distributions

We can define marginal CDFs using the joint CDF by setting one of the values to infinity:

$$\begin{aligned} F(x,\infty) &= P(X \le x, Y \le \infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(x,y)\;dy\;dy \\ &= P(X \le x) = F_X(x) \end{aligned}$$

$$\begin{aligned} F(\infty,y) &= P(X \le \infty, Y \le y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f(x,y)\;dx\;dy \\ &= P(Y \le y) = F_Y(y) \end{aligned}$$

Joint pdf

Similar to the CDF the probability density function follows the same general rules in two dimensions,

Univariate definition:

$f(x) \geq 0$ for all x $\qquad f(x) = \frac{d}{dx}F(x) \qquad \int_{-\infty}^{\infty} f(x)dx = 1$

Bivariate definition:

$$f(x,y) \geq 0 \text{ for all } (x,y)$$

$$f(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x,y)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\; dx\; dy = 1$$

Marginal pdfs

Marginal pdfs are derived by integrating out one of the random variables.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\ dy$$

$$f_Y(x) = \int_{-\infty}^{\infty} f(x,y)\ dx$$

Previously we defined independence in terms of, X and Y are independent if and only if $E(XY) = E(X)E(Y)$.

An equivalent definitions is, X and Y are independent if and only if $f(x,y) = f_X(x)f_Y(y)$.

Probability and Expectation

Univariate definition:

$$P(X \in A) = \int_A f(x)\ dx$$

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

Bivariate definition:

$$P(X \in A, Y \in B) = \int_A \int_B f(x,y)\ dx\ dy$$

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot f(x,y)\ dx\ dy$$

Conditional Expectation

If X and Y are independent random variables then we define the conditional expectation as follows

$$E(X|Y=y) = \sum_{\text{all } y} x f(x|y)\ dx$$

$$E(X|Y=y) = \int_{-\infty}^{\infty} x f(x|y)\ dx$$

We can think of the conditional expectation as being a function of the random variable Y, thereby making $E(X|Y)$ itself a random variable

$$E(E(X|Y)) = E(X)$$

Conditional Variance

Similar to conditional expectation we can also define conditional variance

$$
\begin{aligned}
Var(Y|X) &= E\left[(Y - E(Y|X))^2 \middle| X \right] \\
&= E(Y^2|X) - E(Y|X)^2
\end{aligned}
$$

there is also an equivalent to the law of total probability for expectations, the Law of Total Probability for Variance:

$$Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]$$

Covariance

We have previously discussed Covariance in relation to the variance of the sum of two random variables (Review Lecture 8).

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

Specifically, covariance is defined as

$$\begin{aligned} Cov(X,Y) &= E[(X-E(X))(Y-E(Y))] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

this is a generalization of variance to two random variables and generally measures the degree to which X and Y tend to be large (or small) at the same time or the degree to which one tends to be large while the other is small.

Properties of Covariance

- $Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) - \mu_x\mu_y$
- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, Y) = 0$ if X and Y are independent
- $Cov(X, c) = 0$
- $Cov(X, X) = Var(X)$
- $Cov(aX, bY) = ab\ Cov(X, Y)$
- $Cov(X + a, Y + b) = Cov(X, Y)$
- $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$

Correlation

Since $Cov(X, Y)$ depends on the magnitude of X and Y we would prefer to have a measure of association that is not effected by arbitrary changes in the scales of the random variables.

The most common measure of *linear* association is correlation which is defined as

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

$$-1 < \rho(X, Y) < 1$$

Where the magnitude of the correlation measures the strength of the *linear* association and the sign determines if it is a positive or negative relationship.

General Bivariate Normal

Let $Z_1, Z_2 \sim \mathcal{N}(0,1)$, which we will use to build a general bivariate normal distribution.

$$f(z_1, z_2) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(z_1^2 + z_2^2)\right]$$

We want to transform these unit normal distributions to have the follow arbitrary parameters: $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$

$$\begin{aligned} X &= \sigma_X Z_1 + \mu_X \\ Y &= \sigma_Y[\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_Y \end{aligned}$$

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)}{\sigma_X}\frac{(y-\mu_Y)}{\sigma_Y}\right)\right]$$

General Bivariate Normal - Marginals and Cov

The marginal distributions of X and Y are given by

$$X = \sigma_X Z_1 + \mu_X = \sigma_X \mathcal{N}(0,1) + \mu_X = \mathcal{N}(\mu_X, \sigma_X^2)$$

$$\begin{aligned} Y &= \sigma_Y [\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_Y = \sigma_Y [\rho \mathcal{N}(0,1) + \sqrt{1-\rho^2} \mathcal{N}(0,1)] + \mu_Y \\ &= \sigma_Y [\mathcal{N}(0,\rho^2) + \mathcal{N}(0, 1-\rho^2)] + \mu_Y = \sigma_Y \mathcal{N}(0,1) + \mu_Y \\ &= \mathcal{N}(\mu_Y, \sigma_Y^2) \end{aligned}$$

$$\begin{aligned} Cov(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= E\left[(\sigma_X Z_1 + \mu_X - \mu_X)(\sigma_Y [\rho Z_1 + \sqrt{1-\rho^2} Z_2] + \mu_Y - \mu_Y)\right] \\ &= E\left[(\sigma_X Z_1)(\sigma_Y [\rho Z_1 + \sqrt{1-\rho^2} Z_2])\right] = \sigma_X \sigma_Y E\left[\rho Z_1^2 + \sqrt{1-\rho^2} Z_1 Z_2\right] \\ &= \sigma_X \sigma_Y \rho E[Z_1^2] = \sigma_X \sigma_Y \rho \end{aligned}$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \rho$$

Conditional Expectation of the Bivariate Normal

Using $X = \mu_X + \sigma_X Z_1$ and $Y = \mu_Y + \sigma_Y[\rho Z_1 + (1-\rho^2)^{1/2} Z_2]$ where $Z_1, Z_2 \sim \mathcal{N}(0,1)$ we can find $E(Y|X)$.

$$\begin{aligned} E[Y|X=x] &= E\left[\mu_Y + \sigma_Y\left(\rho Z_1 + (1-\rho^2)^{1/2} Z_2\right)\middle| X = x\right] \\ &= E\left[\mu_Y + \sigma_Y\left(\rho \frac{x-\mu_X}{\sigma_X} + (1-\rho^2)^{1/2} Z_2\right)\middle| X = x\right] \\ &= \mu_Y + \sigma_Y\left(\rho \frac{x-\mu_X}{\sigma_X} + (1-\rho^2)^{1/2} E[Z_2|X=x]\right) \\ &= \mu_Y + \sigma_Y \rho \left(\frac{x-\mu_X}{\sigma_X}\right) \end{aligned}$$

By symmetry,

$$E[X|Y=y] = \mu_X + \sigma_X \rho \left(\frac{y-\mu_Y}{\sigma_Y}\right)$$

Conditional Variance of the Bivariate Normal

Using $X = \mu_X + \sigma_X Z_1$ and $Y = \mu_Y + \sigma_Y [\rho Z_1 + (1-\rho^2)^{1/2} Z_2]$ where $Z_1, Z_2 \sim \mathcal{N}(0,1)$ we can find $Var(Y|X)$.

Conditional Variance of the Bivariate Normal

Using $X = \mu_X + \sigma_X Z_1$ and $Y = \mu_Y + \sigma_Y[\rho Z_1 + (1-\rho^2)^{1/2} Z_2]$ where $Z_1, Z_2 \sim \mathcal{N}(0,1)$ we can find $Var(Y|X)$.

$$
\begin{aligned}
Var[Y|X=x] &= Var\left[\mu_Y + \sigma_Y\left(\rho Z_1 + (1-\rho^2)^{1/2} Z_2\right)\middle| X = x\right] \\
&= Var\left[\mu_Y + \sigma_Y\left(\rho \frac{x-\mu_X}{\sigma_X} + (1-\rho^2)^{1/2} Z_2\right)\middle| X = x\right] \\
&= Var[\sigma_Y (1-\rho^2) Z_2 | X = x] \\
&= \sigma_Y^2 (1-\rho^2)
\end{aligned}
$$

By symmetry,

$$
Var[X|Y=y] = \sigma_X^2 (1-\rho^2)
$$