Lecture 9: Review

Sta230 / Mth230

Colin Rundel

February 19, 2012

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:

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

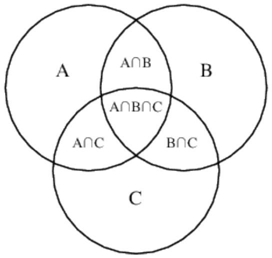
Inclusion-Exclusion:

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

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:

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

Useful Approximations

Log:

$$\log(1+x) \approx x$$
$$\log(1-x) \approx -x$$

Stirling's:

$$k! \approx \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$$

Taylor Series

For any* function $f(x)$ we can rewrite it using

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$

Often each progressive term gets smaller, so we can approximate the function with only the first several terms

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a)$$

.

Useful Taylor Series

$$\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \cdots$$
$$\log(1-x) = -x - x^2/2 - x^3/3 - x^4/4 + \cdots$$

$$\exp(x) = 1 + x + x^2/2 + x^3/6 + x^4/24 + \cdots$$

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, $X \sim \text{Binom}(n,p)$

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

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

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, $Y \sim \text{Geo}(p)$

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

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

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

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

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

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

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

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

Standardizing Normal Distributions

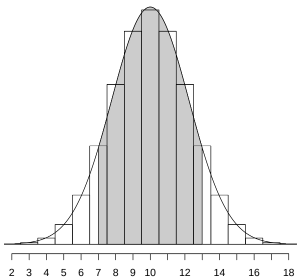
Let X be a normally distributed random variable with mean μ and variance σ^2 then we define the random variable Z such that

$$Z = \left(\frac{X-\mu}{\sigma}\right)$$
$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{E(X)-\mu}{\sigma} = 0$$
$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{Var(X)}{\sigma^2} = 1$$

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

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:

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

Multinomial Distribution

Let $X_1, X_2, \cdots, X_k$ be the k random variables that reflect the number of outcomes belonging to category k in n trials with the probability of success for category k being p_k, $X_1, \cdots, X_k \sim \text{Multinom}(n, p_1, \cdots, p_k)$

$$\begin{aligned} P(X_1 = x_1, \cdots, X_k = x_k) &= f(x_1, \cdots, x_k | n, p_1, \cdots, p_k) \\ &= \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \end{aligned}$$
$$\text{where } \sum_{i=1}^k x_i = n \text{ and } \sum_{i=1}^k p_i = 1$$

$$E(X_i) = np_i$$
$$Var(X_i) = np_i(1-p_i)$$
$$Cov(X_i, X_j) = -np_ip_j$$

Functions of Random Variables

Expected Value:

- Discrete Random Variable:

$$E[X] = \sum_{\text{all } x} xP(X = x)$$

- Continous Random Variable:

$$E[X] = \int_{\text{all } x} xP(X = x)dx$$

Variance:

$$Var(X) = E\left[(X - E(X))^2\right] = E(X - \mu)^2 = E(X)^2 - \mu^2$$

Covariance:

$$Var(X) = E\left[(X - E(X))(Y - E(Y))\right] = E[(X - \mu_x)(Y - \mu_y)]$$

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

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

Properties of Covariance

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

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:

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

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

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 $t = 0$.

$$M_X(0) = E[e^0] = 1 = \mu'_0$$

$$M'_X(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt} e^{tX}\right] = E[Xe^{tX}]$$

$$M'_X(0) = E\left[Xe^0\right] = E[X] = \mu'_1$$

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

$$M''_X(0) = E\left[X^2 e^0\right] = E\left[X^2\right] = \mu'_2$$

1.R.2

A box contains one white ball and one black ball. A ball is drawn at random, then replaced in the box with an additional ball of the same color. Then a second ball is drawn at random from the three balls in the box. What is the probability that the first ball drawn was white given that at least one of the two balls drawn was white.

1.R.6

Show that if A and B are independent then so are A^c and B, A and B^c, and A^c and B^c.

1.R.11

A hat contains n coins, f of which are fair, and b of which are biased to land heads with probability $2/3$. A coin is drawn from the hat and tossed twice. The first time it lands heads and the second time it lands tails. Given this information, what is the probability that it is a fair coin?

2.R.12

In poker, a hand containing face values of the form (x, x, y, z, w) is called one pair.

(a) If I deal a poker hand, what is the probability that I get one pair?

(b) I keep dealing independent poker hands. Write an expression for the probability that I get my 150th 'one pair' on or after the 400th deal.

(c) Approximately what is the value of the probability in (b)?

3.R.13

Let X and Y be independent random variables with $E(X) = E(Y) = \mu$, $Var(X) = Var(Y) = \sigma^2$. Show that $Var(XY) = \sigma^2(2\mu^2 + \sigma^2)$.

CLT Example

Let $X_1, \ldots, X_{300} \overset{iid}{\sim} \text{Unif}(0,1)$ and $S_{300} = X_1 + \cdots + X_{300}$. Find $P(S_{300} > 160)$.

MGF

If $M_X(t) = 1 - p + pe^t$, find $E(X^3)$

Conditional Expectation Example

You flip three quarters, if X is the number of heads, then the conditional expectation what is $E(X|\text{the first toss is heads})$