

Final Review

Alexander Van Roijen

December 1, 2018

Please complete the following:

1. Lecture 1. Basics of Probability

(a) **Definition (Sample space and event).** The sample space S of an experiment is the set of all possible outcomes of the experiment. An event A is a subset of the sample space S , and we say that A occurred if the actual outcome is in A .

(b) **Definition (General definition of probability).** A probability space consists of a sample space S and a probability function $P(\cdot)$ which takes an event $A \subset S$ as input and returns $P(A)$, a real number between 0 and 1, as output. The probability function must satisfy the following axioms: $P(\emptyset) = 0$, $P(S) = 1$ and for a union of disjoint events, we get $P(A_1 \cup \dots A_n) = P(A_1) + \dots P(A_n)$

(c) **Theorem. Properties of probability.** A probability function has the following properties, for any events A and B .

i. $P(A^c) = 1 - P(A)$

ii. if $A \subset B$ $P(A) \leq P(B)$

iii. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ which can be extended

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

(d) **Definition (Conditional probability).** If A and B are events with $P(B) > 0$, then the conditional probability of A given B , denoted by $P(A|B)$, is defined as: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ Further note that all probabilities are in fact conditional. We like to think of $P(A)$ as our prior beliefs of an event, and $P(A|B)$ as our posterior, or what we think it is given something is already known.

(e) **Theorem.** For any events A_1, \dots, A_n with positive probabilities,

$$P(A_1, \dots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \dots P(A_n|A_1, \dots, A_{n-1})$$

(f) **Theorem (Bayes rule).** $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

(g) **Theorem (Law of total probability (LOTP)).** Let A_1, \dots, A_n be a partition of the sample space S with $P(A_i) > 0$ for all i . Then

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

When we condition on an event E , we update our beliefs to be consistent with this

knowledge, effectively putting ourselves in a universe where we know that E occurred. Within our new universe, however, the laws of probability operate just as before. Conditional probability satisfies all the properties of probability!

- (h) **Theorem (Bayes rule with extra conditioning).** Provided that $P(A \cap E) > 0$ and $P(B \cap E) > 0$, we have $P(A|B, E) = \frac{P(B|A, E)P(A|E)}{P(B|E)}$
 - (i) **Theorem (Law of total probability (LOTP) with extra conditioning).** Let $A_1 \dots A_n$ be a partition of the sample space S with $P(A_i|E) > 0$ for all i. Then $P(B|E) = \sum_{i=1}^n P(B|A_i, E)P(A_i|E)$
 - (j) **Definition (Independence of two events).** Events A and B are independent if $P(A \cap B) = P(A)P(B)$. If $P(A) > 0$ and $P(B) > 0$, then this is equivalent with $P(A \cap B) = P(A)$, and also equivalent with $P(B \cap A) = P(B)$. Independence is a symmetric relation.
 - (k) **Proposition.** If A and B are independent, then A^c and B are independent, A^c and B^c are independent, and A and B^c are independent.
 - (l) **Definition (Independence of three events).** Events A, B and C are said to be independent if all of the following relations hold:

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

$$P(A \cap C) = P(A)P(C);$$

$$P(B \cap C) = P(B)P(C);$$

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$
 - (m) **Definition (Conditional independence).** Events A and B are said to be conditionally independent given E if $P(A \cap B|E) = P(A|E)P(B|E)$.
 - (n) Problems shown: Monty Hall, and Positive test of conditionitis and bayes rule
2. (a) **Definition (Random variable).** Given an experiment with sample space S, a random variable is a function from the sample space S to the real numbers R. It is common, but not required, to denote random variables by capital letters. $P(X=x) = P(X=X(s))$
- (b) Discrete PMFs are non negative, and sum to one over their support.

- (c) **Definition (Bernoulli distribution).** An random variable X is said to have a Bernoulli distribution with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$. We write this as $X \sim \text{Bern}(p)$.
- (d) **Theorem.** Let $X \sim \text{Bin}(n, p)$, and $q = 1 - p$ (often taken to denote the failure of a Bernoulli trial). Then $n - X \sim \text{Bin}(n, q)$.
- (e) **Theorem (Hypergeometric PMF).** Consider an urn with w white balls and b blacks balls. We draw n balls out of the urn at random without replacement such that all the $\binom{w+b}{n}$ samples are equally likely. Let X be the number of white balls in the sample. Then X is said to have the Hypergeometric distribution with parameters w, b and n : $X \sim \text{HGeom}(w, b, n)$. Then the PMF of X is
- $$P(X = k) = \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}}$$
- (f) **Theorem.** If $X \sim \text{HGeom}(w, b, n)$ and $Y \sim \text{HGeom}(n, w + b - n, w)$, then X and Y have the same distribution.
- (g) **Theorem.** If $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, and X is independent of Y , then the conditional distribution of $X|X + Y = r \sim \text{hgeom}(n, m, r)$.
- (h) **Theorem (Binomial as a limiting case of the Hypergeometric).** If $X \sim \text{HGeom}(w, b, n)$ and $N = w + b$ approaches infinity such that $p = w/(w + b)$ remains fixed, then the PMF of X converges to the $\text{Bin}(n, p)$ PMF.
- (i) **Theorem (PMF of $g(X)$).** Let X be a discrete random variable and $g: R \rightarrow R$. Then the support of $g(X)$ is the set of all y such that $g(x) = y$ for at least one x in the support of X , and the PMF of $g(X)$ is $P(g(X) = Y) = \sum_{x: g(x)=y} P(X = x)$
- (j) **Definition (Independence of two random variables).** Random variables X and Y are said to be independent if $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ for all $x, y \in R$. In the discrete case, this is equivalent to the condition $P(X = x, Y = y) = P(X = x)P(Y = y)$; for all x in the support of X and all y in the support of Y .

3. chapter 3: Expectation

- (a) **Proposition (Monotonicity of expectation).** Let X and Y be random variables such that $X \geq Y$ with probability 1. Then $E(X) \geq E(Y)$, with equality holding if and only if $X = Y$ with probability 1.

- (b) A hypergeometric can be considered a sum of bernoulli random variables, but with their probabilities conditioned on the previous iteration. However, when calculating the expectation, each of the bernoullis are equally like to be picked first, so the expectation is $n * p$ where $p = w/w+b$

(c) **Theorem. Properties of Indicator Random Variables**

1. $I_A^k = I_A$
2. $I_{A^c} = 1 - I_A$
3. $I_{A \cap B} = I_A I_B$
4. $I_{A \cup B} = I_A + I_B - I_{A \cap B}$

(d) Inclusion Exclusion

$$P(A_1 \cup \dots A_n) = \sum_i P(A_i) - \sum_{j>i} P(A_i \cap A_j) + \dots (-1)^n P(A_1 \dots A_n)$$

- (e) Negative Binomial: Number of failures till a fixed number of successes. $\Rightarrow NBin(r, p)P(X = x) = \binom{x+r-1}{r-1} p^r q^x$

- (f) **Poisson** Note $\sum_{k=1}^{\infty} \frac{x^{k-1}}{(k-1)!} = e^x$ Useful for determining expectations and such. its a kernel

- (g) **Theorem. Sum of independent poisson is poisson** if $X \sim Pois(\lambda_1)$ & $Y \sim Pois(\lambda_2)$ $X + Y \sim Pois(\lambda_1 + \lambda_2)$

- (h) **Theorem. Poisson given a sum of Poissons is Binomial** $X \sim Pois(\lambda_1)$ & $Y \sim Pois(\lambda_2)$ $PoisX = k | X + Y = n \sim Binom(n, \lambda_1/(\lambda_1 + \lambda_2))$

4. Chapter 4: Continuous

- (a) **Universality of the Uniform.** This is useful as it can simulate a RV with any discrete Distribution from the Unif.

ex

$$F(x) = \frac{e^x}{1+e^x} \quad F^{-1}(u) = \log\left(\frac{u}{1-u}\right) \sim \text{logistic with } U \sim Unif(0, 1)$$

(b) **Normal distribution**

Note that the normal is symmetric, meaning tail areas are symmetric, and in general

$$\phi(z) = \phi(-z)$$

- (c) **exponential distribution** This is the continuous version of the geometric distribution. $f(x) = \lambda e^{-\lambda x} x > 0$

- (d) **Generalize exponential** $Y = \frac{X}{\lambda} \sim \text{Expo}(\lambda)$ with $X \sim \text{Expo}(1)$
- (e) see notes for further elaboration on this section and bayesian view point of statistics

5. Chapter 5: Moments

- (a) **Definition (Median).** We say that c is a median of a random variable X if $P(X \leq c) \geq .5$ and $P(X \geq c) \geq .5$
- (b) **Definition (Mode).** For a discrete random variable X , we say that c is a mode of X if it maximizes the PMF: $P(X = c) \geq P(X = x)$ for all x . For a continuous random variable X with PDF f , we say that c is the mode if it maximizes the PDF: $f(c) \geq f(x)$ for all x .

Note we can have multiple modes and medians but not means!

- (c) **Theorem.** Let X be a random variable with mean μ , and let m be the median of X . The value of c that minimizes the mean squared error $E(X - c)^2$ is $c = \mu$. A value of c that minimizes the mean absolute error $E|X - c|$ is $c = m$.
- (d) **Definition (Symmetry of a random variable).** We say that a random variable X has a symmetric distribution about μ if $X - \mu$ has the same distribution as $\mu - X$. The number μ must be the mean $E(X)$ if it exists, and must also be a median of the distribution of X .
- (e) **Proposition (Symmetry in terms of the PDF).** Let X be a continuous random variable with PDF f . Then X is symmetric about μ if and only if $f(x) = f(2\mu - x)$ for all x .
- (f) **Definition (Kinds of moments).** Let X be a random variable with mean μ and variance σ^2 . For any positive integer n , the n th moment of X is $E(X^n)$, the n th central moment is $E[(X - \mu)^n]$, and the n th standardized moment is $E((\frac{X - \mu}{\sigma})^n)$. The mean is the first moment and the variance is the second central moment
- (g) **Definition (Skewness).** The skewness of a random variable X with mean μ and variance σ^2 is the third standardized moment of X

$$\text{Skew}(X) = E[(\frac{X - \mu}{\sigma})^3]$$
- (h) **Proposition (Odd central moments of a symmetric distribution).** Let X be symmetric about its mean μ . Then for any odd number m , the m th central moment

$E(X - \mu)^m$ is 0 if it exists.

- (i) **kurtosis** This represents how different a distribution's pdf is, large kurtosis means sharp peak in the middle, heavy tails, and low shoulders.

- (j) **Definition (Sample moments).** Let X_1, X_2, \dots, X_n be i.i.d. random variables. The k th sample moment is the random variable

$$M_k = \frac{1}{n} \sum_{j=1}^n X_j^k$$

and $E[M_k] = E\left[\frac{1}{n} \sum_{j=1}^n X_j^k\right] = E[X_1^k]$. Hence it is an unbiased estimator of the true k th moment

- (k) $Var(\bar{X}_n) = \frac{\sigma^2}{n}$

- (l) **Sample variance** $= S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2$

- (m) **Example: Bernoulli MGF** For $X \sim \text{Bern}(p)$, the MGF is $M(t) = E(e^{tX}) = e^{t0}P(X=0) + e^{t1}P(X=1) = pe^t + q$ where $q = 1 - p$. Since $M(t)$ is finite for any $t \in \mathbb{R}$, the MGF of a Bernoulli random variable is defined on the entire real line.

- (n) **Example: Geometric MGF** For $X \sim \text{Geom}(p)$, the MGF is $M(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} P(X=k) = \sum_{k=0}^{\infty} e^{tk} q^k p = \frac{p}{1 - qe^t}$ for $qe^t < 1$ or, equivalently, for $(-\infty, \log(1/q))$,

- (o) **Theorem (Moments via derivatives of MGFs).** Given the MGF $M(t)$ of a random variable X , we can get the n th moment of X by evaluating its n th derivative of the MGF at 0 $E(X^n) = M^{(n)}(0)$

- (p) **Theorem (MGF determines the distribution).** The MGF of a random variable determines its distribution: if two random variables have the same MGF, they must have the same distribution.

- (q) **Theorem (MGF of a sum of independent random variables).** If X and Y are independent, then the MGF of $X + Y$ is the product of the individual MGFs: $M_{X+Y}(t) = M_X(t)M_Y(t)$

Using this result, we can obtain the MGFs of the Binomial and Negative Binomial, which are sums of independent Bernoulli and Geometric i.i.d. random variables.

Example: Binomial MGF The MGF of a $\text{Bern}(p)$ random variable is $pe^t + q$, so the MGF of a $\text{Bin}(n, p)$ random variable is $M(t) = (pe^t + q)^n$:

Happy holidays!