Probability Formula Sheet *Haris H.*

Axioms

Axiom 1: $0 \leq P[A]$

It states that the probability (mass) is nonnegative

Axiom 2: $P[S] = 1$

It states that there is a fixed total amount of probability (mass), namely 1 unit.

Axiom 3: If $A \cap B = \emptyset$, then $P[A \cup B] = P[A] + P[B]$

It states that the total probability (mass) in two disjoint objects is the sum of the individual probabilities (masses).

Axiom 4:

If $A_1, A_2, \ldots$ is a sequence of events such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$$

Corollaries	
$P[A^c]=1-P[A]$	$1=P[S]=P[A^c]+P[A]$
$P[A]\leq 1$	$P[\emptyset]=0$
$P[A \cup B] = P[A] + P[B] - P[A \cap B]$	If $A \subset B$, then $P[A]\leq P[B]$
$P[A \cup B] \leq P[A] + P[B]$	

Computing probabilities by counting methods

- **Sampling with Replacement and with Ordering:**

Number of distinct ordered k-tuples = n^k

- **Sampling without Replacement and with Ordering:**

Number of distinct ordered k-tuples = $n(n-1)\ldots(n-k+1)$

Permutations of *n* Distinct Objects

Number of permutations of n objects = $n(n-1)\ldots(2)(1) \triangleq n!$ (We refer to $n!$ as ***n* factorial**.)

- **Sampling without Replacement and without Ordering**

$$C_k^n\, k! = n(n-1)\ldots(n-k+1)$$

$$C_k^n = \frac{n(n-1)\ldots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \triangleq \binom{n}{k}$$

- **Sampling with Replacement and without Ordering**

$$\binom{n-1+k}{k} = \binom{n-1+k}{n-1}$$

Conditional probability

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

Addition Rule:

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

If A & B are mutually exclusive then; $P[A \cup B] = P[A] + P[B]$

Multiplication Rule:

$$P[A \cap B] = P[A]P[B|A] = P[B]P[A|b]$$

If A & B are independent then; $P[A \cap B] = P[A]P[B]$

Probability Laws

Total probability law:

$P[A] = P[A \cap B_1] + P[A \cap B_2] + \cdots + P[A \cap B_n]$ OR

$P[A] = P[A|B_1]P[B_1] + P[A|B_2]P[B_2] + \cdots + P[A|B_n]P[B_n]$ OR

$P[A] = P[A|B]P[B] + P[A|B^c]P[B^c]$

Bayes' Law

$$P[B_i|A] = \frac{P[B_i \cap A]}{P[A]} = \frac{P[B_i \cap A]}{P[A]} = P[A|B_i]\frac{P[B_i]}{P[A]} = \frac{P[A|B_i]P[B_i]}{\sum_{k=1}^{n} P[A|B_k]P[B_k]}$$

Probability definition:

A & B are mutually exclusive if $[A \cap B] = 0$

A & B are independent if $P[A|B] = P[A]$ & $P[B|A] = P[B]$ & $P[A \cap B] = P[A]P[B]$

Discrete Probability Distributions

Expected value or mean of a discrete random variable *X*:

$$m_X = E[X] = \sum_{x \in S_X} x p_X(x) = \sum_k x_k p_X(x_k)$$

Variance of the random variable *X*

$$\sigma_X^2 = VAR[X] = E[(X - m_X)^2] = \sum_k (x_k - m_X)^2 p_X(x_k) = E[X^2] - m_X^2$$

Standard deviation of the random variable *X*:

$$\sigma = STD[X] = VAR[X]^{1/2}$$

TNT

X	X counts	P_X		Values of x		E[X]		VAR[X]
Bernoulli	Equals one if the event *A* occurs, and zero otherwise.	$P_0=1-p,\ p_1=p$		0,1		p		P(1-p)
Binomial	Numbers of successes in fixed n trials	$\binom{n}{k}p^k(1-p)^{n-k}$		0,1,...,n		np		np(1-p)
Geometric	Number of trials up through 1st success	$p(1-p)^k$	$p(1-p)^{k-1}$	0, 1, ..	1, 2, ...	$\frac{1-p}{p}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Uniform	outcomes are equaly likely	$\frac{1}{L}$		1,2,..,L		$\frac{L+1}{2}$		$\frac{L^2-1}{12}$
Poisson	number of events that occur in fixed time period	$\frac{\alpha^k}{k!}e^{-\alpha}$, $\alpha > 0$		0,1,2		α		α

Probability Cheatsheet v1.1

Compiled by William Chen (http://wzchen.com) with contributions from Sebastian Chiu, Yuan Jiang, Yuqi Hou, and Jessy Hwang. Material based off of Joe Blitzstein's (@stat110) lectures (http://stat110.net) and Blitzstein/Hwang's Intro to Probability textbook (http://bit.ly/introprobability).

Last Updated February 28, 2015

Counting

Multiplication Rule - Let's say we have a compound experiment (an experiment with multiple components). If the 1st component has n_1 possible outcomes, the 2nd component has n_2 possible outcomes, and the rth component has n_r possible outcomes, then overall there are $n_1 n_2 \dots n_r$ possibilities for the whole experiment.

Sampling Table - The sampling tables describes the different ways to take a sample of size k out of a population of size n. The column names denote whether order matters or not.

	Matters	Not Matter
With Replacement	n^k	$\binom{n+k-1}{k}$
Without Replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Naïve Definition of Probability - *If the likelihood of each outcome is equal*, the probability of any event happening is:

$$P(\text{Event}) = \frac{\text{number of favorable outcomes}}{\text{number of outcomes}}$$

Probability and Thinking Conditionally

Independence

Independent Events - **A** and **B** are independent if knowing one gives you no information about the other. **A** and **B** are independent if and only if one of the following equivalent statements hold:

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

Conditional Independence - **A** and **B** are conditionally independent given **C** if: $P(\mathbf{A} \cap \mathbf{B}|\mathbf{C}) = P(\mathbf{A}|\mathbf{C})P(\mathbf{B}|\mathbf{C})$. Conditional independence does not imply independence, and independence does not imply conditional independence.

Unions, Intersections, and Complements

De Morgan's Laws - Gives a useful relation that can make calculating probabilities of unions easier by relating them to intersections, and vice versa. De Morgan's Law says that the complement is distributive as long as you flip the sign in the middle.

$$(\mathbf{A} \cup \mathbf{B})^c \equiv \mathbf{A}^c \cap \mathbf{B}^c$$
$$(\mathbf{A} \cap \mathbf{B})^c \equiv \mathbf{A}^c \cup \mathbf{B}^c$$

Joint, Marginal, and Conditional Probabilities

Joint Probability - $P(\mathbf{A} \cap \mathbf{B})$ or $P(\mathbf{A}, \mathbf{B})$ - Probability of **A** and **B**

Marginal (Unconditional) Probability - $P(\mathbf{A})$ - Probability of **A**

Conditional Probability - $P(\mathbf{A}|\mathbf{B})$ - Probability of **A** given **B** occurred.

Conditional Probability is Probability - $P(\mathbf{A}|\mathbf{B})$ is a probability as well, restricting the sample space to **B** instead of Ω. Any theorem that holds for probability also holds for conditional probability.

Simpson's Paradox

$$P(A \mid B, C) < P(A \mid B^c, C) \text{ and } P(A \mid B, C^c) < P(A \mid B^c, C^c)$$
$$\text{yet still, } P(A \mid B) > P(A \mid B^c)$$

Bayes' Rule and Law of Total Probability

Law of Total Probability with partitioning set $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \dots \mathbf{B}_n$ and with extra conditioning (just add C!)

$$P(\mathbf{A}) = P(\mathbf{A}|\mathbf{B}_1)P(\mathbf{B}_1) + P(\mathbf{A}|\mathbf{B}_2)P(\mathbf{B}_2) + \dots P(\mathbf{A}|\mathbf{B}_n)P(\mathbf{B}_n)$$
$$P(\mathbf{A}) = P(\mathbf{A} \cap \mathbf{B}_1) + P(\mathbf{A} \cap \mathbf{B}_2) + \dots P(\mathbf{A} \cap \mathbf{B}_n)$$
$$P(\mathbf{A}|\mathbf{C}) = P(\mathbf{A}|\mathbf{B}_1, \mathbf{C})P(\mathbf{B}_1|\mathbf{C}) + \dots P(\mathbf{A}|\mathbf{B}_n, \mathbf{C})P(\mathbf{B}_n|\mathbf{C})$$
$$P(\mathbf{A}|\mathbf{C}) = P(\mathbf{A} \cap \mathbf{B}_1|\mathbf{C}) + P(\mathbf{A} \cap \mathbf{B}_2|\mathbf{C}) + \dots P(\mathbf{A} \cap \mathbf{B}_n|\mathbf{C})$$

Law of Total Probability with **B** and $\mathbf{B}^c$ (special case of a partitioning set), and with extra conditioning (just add C!)

$$P(\mathbf{A}) = P(\mathbf{A}|\mathbf{B})P(\mathbf{B}) + P(\mathbf{A}|\mathbf{B}^c)P(\mathbf{B}^c)$$
$$P(\mathbf{A}) = P(\mathbf{A} \cap \mathbf{B}) + P(\mathbf{A} \cap \mathbf{B}^c)$$
$$P(\mathbf{A}|\mathbf{C}) = P(\mathbf{A}|\mathbf{B}, \mathbf{C})P(\mathbf{B}|\mathbf{C}) + P(\mathbf{A}|\mathbf{B}^c, \mathbf{C})P(\mathbf{B}^c|\mathbf{C})$$
$$P(\mathbf{A}|\mathbf{C}) = P(\mathbf{A} \cap \mathbf{B}|\mathbf{C}) + P(\mathbf{A} \cap \mathbf{B}^c|\mathbf{C})$$

Bayes' Rule, and with extra conditioning (just add C!)

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

Odds Form of Bayes' Rule, and with extra conditioning (just add C!)

$$\frac{P(\mathbf{A}|\mathbf{B})}{P(\mathbf{A}^c|\mathbf{B})} = \frac{P(\mathbf{B}|\mathbf{A})}{P(\mathbf{B}|\mathbf{A}^c)} \frac{P(\mathbf{A})}{P(\mathbf{A}^c)}$$
$$\frac{P(\mathbf{A}|\mathbf{B}, \mathbf{C})}{P(\mathbf{A}^c|\mathbf{B}, \mathbf{C})} = \frac{P(\mathbf{B}|\mathbf{A}, \mathbf{C})}{P(\mathbf{B}|\mathbf{A}^c, \mathbf{C})} \frac{P(\mathbf{A}|\mathbf{C})}{P(\mathbf{A}^c|\mathbf{C})}$$

Random Variables and their Distributions

PMF, CDF, and Independence

Probability Mass Function (PMF) (Discrete Only) gives the probability that a random variable takes on the value X.

$$P_X(x) = P(X = x)$$

Cumulative Distribution Function (CDF) gives the probability that a random variable takes on the value x or less

$$F_X(x_0) = P(X \leq x_0)$$

Independence - Intuitively, two random variables are independent if knowing one gives you no information about the other. X and Y are independent if for ALL values of x and y:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

Expected Value and Indicators

Distributions

Probability Mass Function (PMF) (Discrete Only) is a function that takes in the value x, and gives the probability that a random variable takes on the value x. The PMF is a positive-valued function, and $\sum_x P(X = x) = 1$

$$P_X(x) = P(X = x)$$

Cumulative Distribution Function (CDF) is a function that takes in the value x, and gives the probability that a random variable takes on the value at most x.

$$F(x) = P(X \leq x)$$

Expected Value, Linearity, and Symmetry

Expected Value (aka *mean*, *expectation*, or *average*) can be thought of as the "weighted average" of the possible outcomes of our random variable. Mathematically, if $x_1, x_2, x_3, \dots$ are all of the possible values that X can take, the expected value of X can be calculated as follows:

$$E(X) = \sum_i x_i P(X = x_i)$$

Note that for *any* X and Y, a and b scaling coefficients and c is our constant, the following property of **Linearity of Expectation** holds:

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

If two Random Variables have the same distribution, even *when they are dependent* by the property of **Symmetry** their expected values are equal.

Conditional Expected Value is calculated like expectation, only conditioned on any event A.

$$E(X|A) = \sum_x xP(X = x|A)$$

Indicator Random Variables

Indicator Random Variables is random variable that takes on either 1 or 0. The indicator is always an indicator of some event. If the event occurs, the indicator is 1, otherwise it is 0. They are useful for many problems that involve counting and expected value.

Distribution $I_A \sim \text{Bern}(p)$ where $p = P(A)$

Fundamental Bridge The expectation of an indicator for A is the probability of the event. $E(I_A) = P(A)$. Notation:

$$I_A = \begin{cases} 1 & \text{A occurs} \\ 0 & \text{A does not occur} \end{cases}$$

Poisson, Continuous RVs, LotUS, UoU

Continuous Random Variables

What's the prob that a CRV is in an interval? Use the CDF (or the PDF, see below). To find the probability that a CRV takes on a value in the interval $[a, b]$, subtract the respective CDFs.

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

Note that for an r.v. with a normal distribution,

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$
$$= \Phi\left(\frac{b - \mu}{\sigma^2}\right) - \Phi\left(\frac{a - \mu}{\sigma^2}\right)$$

What is the Cumulative Density Function (CDF)? It is the following function of x.

$$F(x) = P(X \leq x)$$

What is the Probability Density Function (PDF)? The PDF, $f(x)$, is the derivative of the CDF.

$$F'(x) = f(x)$$

Or alternatively,

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Note that by the fundamental theorem of calculus,

$$F(b) - F(a) = \int_a^b f(x)dx$$

Thus to find the probability that a CRV takes on a value in an interval, you can integrate the PDF, thus finding the area under the density curve.

How do I find the expected value of a CRV? Where in discrete cases you sum over the probabilities, in continuous cases you integrate over the densities.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Law of the Unconscious Statistician (LotUS)

Expected Value of Function of RV Normally, you would find the expected value of X this way:

$$E(X) = \Sigma_x xP(X = x)$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

LotUS states that you can find the expected value of a *function of a random variable* g(X) this way:

$$E(g(X)) = \Sigma_x g(x)P(X = x)$$

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

What's a function of a random variable? A function of a random variable is also a random variable. For example, if X is the number of bikes you see in an hour, then $g(X) = 2X$ could be the number of bike wheels you see in an hour. Both are random variables.

What's the point? You don't need to know the PDF/PMF of $g(X)$ to find its expected value. All you need is the PDF/PMF of X.

Variance, Expectation and Independence, and e^x Taylor Series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

If X and Y are independent, then

$$E(XY) = E(X)E(Y)$$

Universality of Uniform

When you plug any random variable into its own CDF, you get a Uniform[0,1] random variable. When you put a Uniform[0,1] into an inverse CDF, you get the corresponding random variable. For example, let's say that a random variable X has a CDF

$$F(x) = 1 - e^{-x}$$

By the Universality of the the Uniform, if we plug in X into this function then we get a uniformly distributed random variable.

$$F(X) = 1 - e^{-X} \sim U$$

Similarly, since $F(X) \sim U$ then $X \sim F^{-1}(U)$. The key point is that *for any continuous random variable X, we can transform it into a uniform random variable and back by using its CDF.*

Exponential Distribution and MGFs

Can I Have a Moment?

Moment - Moments describe the shape of a distribution. The first three moments, are related to Mean, Variance, and Skewness of a distribution. The k^{th} moment of a random variable X is

$$\mu'_k = E(X^k)$$

What's a moment? Note that

Mean $\mu'_1 = E(X)$

Variance $\mu'_2 = E(X^2) = \text{Var}(X) + (\mu'_1)^2$

Mean, Variance, and other moments (Skewness) can be expressed in terms of the moments of a random variable!

Moment Generating Functions

MGF For any random variable X, this expected value and function of dummy variable t;

$$M_X(t) = E(e^{tX})$$

is the **moment generating function (MGF)** of X if it exists for a finitely-sized interval centered around 0. Note that the MGF is just a function of a dummy variable t.

Why is it called the Moment Generating Function? Because the k^{th} derivative of the moment generating function evaluated 0 is the k^{th} moment of X!

$$\mu'_k = E(X^k) = M_X^{(k)}(0)$$

This is true by Taylor Expansion of e^{tX}

$$M_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} \frac{E(X^k)t^k}{k!} = \sum_{k=0}^{\infty} \frac{\mu'_k t^k}{k!}$$

Or by differentiation under the integral sign and then plugging in $t = 0$

$$M_X^{(k)}(t) = \frac{d^k}{dt^k}E(e^{tX}) = E(\frac{d^k}{dt^k}e^{tX}) = E(X^k e^{tX})$$

$$M_X^{(k)}(0) = E(X^k e^{0X}) = E(X^k) = \mu'_k$$

MGF of linear combinations If we have $Y = aX + c$, then

$$M_Y(t) = E(e^{t(aX+c)}) = e^{ct}E(e^{(at)X}) = e^{ct}M_X(at)$$

Uniqueness of the MGF. *If it exists, the MGF uniquely defines the distribution.* This means that for any two random variables X and Y, they are distributed the same (their CDFs/PDFs are equal) if and only if their MGF's are equal. You can't have different PDFs when you have two random variables that have the same MGF.

Summing Independent R.V.s by Multiplying MGFs. If X and Y are independent, then

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

$$M_{(X+Y)}(t) = M_X(t) \cdot M_Y(t)$$

The MGF of the sum of two random variables is the product of the MGFs of those two random variables.

Joint PDFs and CDFs

Joint Distributions

Review: Joint Probability of events A and B: $P(A \cap B)$

Both the Joint PMF and Joint PDF must be non-negative and sum/integrate to 1. ($\sum_x \sum_y P(X = x, Y = y) = 1$) ($\int_x \int_y f_{X,Y}(x, y) = 1$). Like in the univariate cases, you sum/integrate the PMF/PDF to get the CDF.

Conditional Distributions

Review: By Baye's Rule, $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ Similar conditions apply to conditional distributions of random variables.

For discrete random variables:

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}$$

For continuous random variables:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$$

Hybrid Bayes' Rule

$$f(x|A) = \frac{P(A|X = x)f(x)}{P(A)}$$

Marginal Distributions

Review: Law of Total Probability Says for an event A and partition $B_1, B_2, ... B_n$: $P(A) = \sum_i P(A \cap B_i)$

To find the distribution of one (or more) random variables from a joint distribution, sum or integrate over the irrelevant random variables.

Getting the Marginal PMF from the Joint PMF

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

Getting the Marginal PDF from the Joint PDF

$$f_X(x) = \int_y f_{X,Y}(x, y)dy$$

Independence of Random Variables

Review: A and B are independent if and only if either $P(A \cap B) = P(A)P(B)$ or $P(A|B) = P(A)$.

Similar conditions apply to determine whether random variables are independent - two random variables are independent if their joint distribution function is simply the product of their marginal distributions, or that the a conditional distribution of is the same as its marginal distribution.

In words, random variables X and Y are independent for all x, y, if and only if one of the following hold:

- Joint PMF/PDF/CDFs are the product of the Marginal PMF
- Conditional distribution of X given Y is the same as the marginal distribution of X

Multivariate LotUS

Review: $E(g(X)) = \sum_x g(x)P(X = x)$, or $E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$

For discrete random variables:

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

For continuous random variables:

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dxdy$$

Covariance and Transformations

Covariance and Correlation

Covariance is the two-random-variable equivalent of Variance, defined by the following:

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

Note that

$$\text{Cov}(X, X) = E(XX) - E(X)E(X) = \text{Var}(X)$$

Correlation is a rescaled variant of Covariance that is always between -1 and 1.

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Covariance and Independence - If two random variables are independent, then they are uncorrelated. The inverse is not necessarily true.

$$X \perp\!\!\!\perp Y \longrightarrow \text{Cov}(X, Y) = 0$$

$$X \perp\!\!\!\perp Y \longrightarrow E(XY) = E(X)E(Y)$$

Covariance and Variance - Note that

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$$

Probability Cheat Sheet

Distributions

Unifrom Distribution

`notation`	$U[a,b]$
`cdf`	$\frac{x-a}{b-a}$ for $x \in [a,b]$
`pdf`	$\frac{1}{b-a}$ for $x \in [a,b]$
`expectation`	$\frac{1}{2}(a+b)$
`variance`	$\frac{1}{12}(b-a)^2$
`mgf`	$\frac{e^{tb}-e^{ta}}{t(b-a)}$

`story:` all intervals of the same length on the distribution's support are equally probable.

Gamma Distribution

`notation`	$Gamma(k,\theta)$
`pdf`	$\frac{\theta^k x^{k-1} e^{-\theta x}}{\Gamma(k)} \mathbb{I}_{x>0}$ $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$
`expectation`	$k\theta$
`variance`	$k\theta^2$
`mgf`	$(1-\theta t)^{-k}$ for $t < \frac{1}{\theta}$
`ind. sum`	$\sum_{i=1}^n X_i \sim Gamma\left(\sum_{i=1}^n k_i, \theta\right)$

`story:` the sum of k independent exponentially distributed random variables, each of which has a mean of θ (which is equivalent to a rate parameter of θ^{-1}).

Geometric Distribution

`notation`	$G(p)$
`cdf`	$1-(1-p)^k$ for $k \in \mathbb{N}$
`pmf`	$(1-p)^{k-1} p$ for $k \in \mathbb{N}$
`expectation`	$\frac{1}{p}$
`variance`	$\frac{1-p}{p^2}$
`mgf`	$\frac{pe^t}{1-(1-p)e^t}$

`story:` the number X of Bernoulli trials needed to get one success. Memoryless.

Poisson Distribution

`notation`	$Poisson(\lambda)$
`cdf`	$e^{-\lambda} \sum_{i=0}^k \frac{\lambda^i}{i!}$
`pmf`	$\frac{\lambda^k}{k!} \cdot e^{-\lambda}$ for $k \in \mathbb{N}$
`expectation`	λ
`variance`	λ
`mgf`	$\exp\left(\lambda\left(e^t-1\right)\right)$
`ind. sum`	$\sum_{i=1}^n X_i \sim Poisson\left(\sum_{i=1}^n \lambda_i\right)$

`story:` the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event.

Normal Distribution

`notation`	$N\left(\mu,\sigma^2\right)$
`pdf`	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$
`expectation`	μ
`variance`	σ^2
`mgf`	$\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$
`ind. sum`	$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$

`story:` describes data that cluster around the mean.

Standard Normal Distribution

`notation`	$N(0,1)$
`cdf`	$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$
`pdf`	$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
`expectation`	$\frac{1}{\lambda}$
`variance`	$\frac{1}{\lambda^2}$
`mgf`	$\exp\left(\frac{t^2}{2}\right)$

`story:` normal distribution with $\mu = 0$ and $\sigma = 1$.

Exponential Distribution

`notation`	$exp(\lambda)$
`cdf`	$1 - e^{-\lambda x}$ for $x \geq 0$
`pdf`	$\lambda e^{-\lambda x}$ for $x \geq 0$
`expectation`	$\frac{1}{\lambda}$
`variance`	$\frac{1}{\lambda^2}$
`mgf`	$\frac{\lambda}{\lambda - t}$
`ind. sum`	$\sum_{i=1}^k X_i \sim Gamma(k,\lambda)$
`minimum`	$\sim exp\left(\sum_{i=1}^k \lambda_i\right)$

`story:` the amount of time until some specific event occurs, starting from now, being memoryless.

Binomial Distribution

`notation`	$Bin(n,p)$
`cdf`	$\sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}$
`pmf`	$\binom{n}{i} p^i (1-p)^{n-i}$
`expectation`	np
`variance`	$np(1-p)$
`mgf`	$\left(1-p+pe^t\right)^n$

`story:` the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p.

Basics

Comulative Distribution Function

$F_X(x) = \mathbb{P}(X \leq x)$

Probability Density Function

$F_X(x) = \int_{-\infty}^{\infty} f_X(t)\, dt$

$\int_{-\infty}^{\infty} f_X(t)\, dt = 1$

$f_X(x) = \frac{d}{dx} F_X(x)$

Quantile Function

The function $X^* : [0,1] \to \mathbb{R}$ for which for any $p \in [0,1]$, $F_X\left(X^*(p)^-\right) \leq p \leq F_X\left(X^*(p)\right)$

$F_{X^*} = F_X$

$\mathbb{E}(X^*) = \mathbb{E}(X)$

Expectation

$\mathbb{E}(X) = \int_0^1 X^*(p) dp$

$\mathbb{E}(X) = \int_{-\infty}^0 F_X(t)\, dt + \int_0^\infty (1 - F_X(t))\, dt$

$\mathbb{E}(X) = \int_{-\infty}^\infty x f_X x dx$

$\mathbb{E}(g(X)) = \int_{-\infty}^\infty g(x) f_X x dx$

$\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$

Variance

$\mathrm{Var}(X) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2$

$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$

$\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$

Standard Deviation

$\sigma(X) = \sqrt{\mathrm{Var}(X)}$

Covariance

$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

$\mathrm{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(x))(Y - \mathbb{E}(Y)))$

$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$

Correlation Coefficient

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

Moment Generating Function

$M_X(t) = \mathbb{E}\left(e^{tX}\right)$

$\mathbb{E}(X^n) = M_X^{(n)}(0)$

$M_{aX+b}(t) = e^{tb} M_{aX}(t)$

Joint Distribution

$\mathbb{P}_{X,Y}(B) = \mathbb{P}((X,Y) \in B)$
$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$

Joint Density

$$\mathbb{P}_{X,Y}(B) = \iint_B f_{X,Y}(s,t)\, ds dt$$
$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)\, dt ds$$
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(s,t)\, ds dt = 1$$

Marginal Distributions

$\mathbb{P}_X(B) = \mathbb{P}_{X,Y}(B \times \mathbb{R})$
$\mathbb{P}_Y(B) = \mathbb{P}_{X,Y}(\mathbb{R} \times Y)$
$$F_X(a) = \int_{-\infty}^{a} \int_{-\infty}^{\infty} f_{X,Y}(s,t)\, dt ds$$
$$F_Y(b) = \int_{-\infty}^{b} \int_{-\infty}^{\infty} f_{X,Y}(s,t)\, ds dt$$

Marginal Densities

$$f_X(s) = \int_{-\infty}^{\infty} f_{X,Y}(s,t) dt$$
$$f_Y(t) = \int_{-\infty}^{\infty} f_{X,Y}(s,t) ds$$

Joint Expectation

$$\mathbb{E}(\varphi(X,Y)) = \iint_{\mathbb{R}^2} \varphi(x,y)\, f_{X,Y}(x,y)\, dx dy$$

Independent r.v.

$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\, \mathbb{P}(Y \leq y)$
$F_{X,Y}(x,y) = F_X(x)\, F_Y(y)$
$f_{X,Y}(s,t) = f_X(s)\, f_Y(t)$
$\mathbb{E}(XY) = \mathbb{E}(X)\, \mathbb{E}(Y)$
$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
Independent events:
$\mathbb{P}(A \cap B) = \mathbb{P}(A)\, \mathbb{P}(B)$

Conditional Probability

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

`bayes` $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(B \mid A)\, \mathbb{P}(A)}{\mathbb{P}(B)}$

Conditional Density

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$
$$f_{X|Y=n}(x) = \frac{f_X(x)\, \mathbb{P}(Y = n \mid X = x)}{\mathbb{P}(Y = n)}$$
$$F_{X|Y=y} = \int_{-\infty}^{x} f_{X|Y=y}(t)\, dt$$

Conditional Expectation

$$\mathbb{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y=y}(x)\, dx$$
$\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X)$
$\mathbb{P}(Y = n) = \mathbb{E}(\mathbb{I}_{Y=n}) = \mathbb{E}(\mathbb{E}(\mathbb{I}_{Y=n} \mid X))$

Sequences and Limits

$$\limsup A_n = \{A_n \text{ i.o.}\} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$$
$$\liminf A_n = \{A_n \text{ eventually}\} = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n$$
$\liminf A_n \subseteq \limsup A_n$
$(\limsup A_n)^c = \liminf A_n^c$
$(\liminf A_n)^c = \limsup A_n^c$
$$\mathbb{P}(\limsup A_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n\right)$$
$$\mathbb{P}(\liminf A_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{n=m}^{\infty} A_n\right)$$

Borel-Cantelli Lemma

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \Rightarrow \mathbb{P}(\limsup A_n) = 0$$
And if A_n are independent:
$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \Rightarrow \mathbb{P}(\limsup A_n) = 1$$

Convergence

Convergence in Probability

`notation` $X_n \xrightarrow{p} X$

`meaning` $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$

Convergence in Distribution

`notation` $X_n \xrightarrow{D} X$

`meaning` $\lim_{n\to\infty} F_n(x) = F(x)$

Almost Sure Convergence

`notation` $X_n \xrightarrow{a.s.} X$

`meaning` $\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$

Criteria for a.s. Convergence

- $\forall \varepsilon \exists N \forall n > N : \mathbb{P}(|X_n - X| < \varepsilon) > 1 - \varepsilon$
- $\forall \varepsilon \mathbb{P}(\limsup(|X_n - X| > \varepsilon)) = 0$
- $\forall \varepsilon \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$ (by B.C.)

Convergence in L_p

`notation` $X_n \xrightarrow{L_p} X$

`meaning` $\lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = 0$

Relationships

$$\begin{array}{ccccc} \xrightarrow{L_q} & \underset{q>p\geq 1}{\Rightarrow} & \xrightarrow{L_p} & & \\ & & \Downarrow & & \\ \xrightarrow{a.s.} & \Rightarrow & \xrightarrow{p} & \Rightarrow & \xrightarrow{D} \end{array}$$

If $X_n \xrightarrow{D} c$ then $X_n \xrightarrow{p} c$
If $X_n \xrightarrow{p} X$ then there exists a subsequence n_k s.t. $X_{n_k} \xrightarrow{a.s.} X$

Laws of Large Numbers

If X_i are i.i.d. r.v.,

`weak law` $\overline{X_n} \xrightarrow{p} \mathbb{E}(X_1)$

`strong law` $\overline{X_n} \xrightarrow{a.s.} \mathbb{E}(X_1)$

Central Limit Theorem

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$
If $t_n \to t$, then
$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t_n\right) \to \Phi(t)$$

Inequalities

Markov's inequality

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}(|X|)}{t}$$

Chebyshev's inequality

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

Chernoff's inequality

Let $X \sim Bin(n,p)$; then:
$\mathbb{P}(X - \mathbb{E}(X) > t\sigma(X)) < e^{-t^2/2}$
Simpler result; for every X:
$\mathbb{P}(X \geq a) \leq M_X(t)\, e^{-ta}$

Jensen's inequality

for φ a convex function, $\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X))$

Miscellaneous

$$\mathbb{E}(Y) < \infty \iff \sum_{n=0}^{\infty} \mathbb{P}(Y > n) < \infty \ (Y \geq 0)$$
$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n) \ (X \in \mathbb{N})$$
$X \sim U(0,1) \iff -\ln X \sim exp(1)$

Convolution

For ind. X, Y, $Z = X + Y$:
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(s)\, f_Y(z-s)\, ds$$

Kolmogorov's 0-1 Law

If A is in the tail σ-algebra $\mathcal{F}^t$, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$

Ugly Stuff

cdf of Gamma distribution:
$$\int_0^t \frac{\theta^k x^{k-1} e^{-\theta k}}{(k-1)!} dx$$

This cheatsheet was made by Peleg Michaeli in January 2010, using LaTeX.
version: 1.01
comments: peleg.michaeli@math.tau.ac.il