Stat 110 Cheatsheet

Go through first and put a box with a blank for the answer. Then set up the problem. Rewrite variables *and* what they mean, especially for questions further from the top. Don't just use formulas, also name them!

Math

$$\frac{d}{dx}e^{-x^2a} = 2xae^{-ax^2}$$

Logs Always take the log on the r.v., not in the probability statement!

$$log(a*b) = log(a) + log(b)$$
$$log_2(q^k) = k * log_2(q)$$

Taylor Series

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

Geometric Series

$$\sum_{n=0}^{n-1} ar^k = a\left(\frac{1-r^n}{1-r}\right)$$

Probability and Counting

Union - $\mathbf{A} \cup \mathbf{B}$ means **A** *or* **B** and you **ADD** up the probabilities

Intersection - $\mathbf{A} \cap \mathbf{B}$ means **A** *and* **B** and you **MULTIPLY** the probabilities

Principle of Inclusion-Exclusion

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \cap \mathbf{B})$$

De Morgan's Laws - De Morgan's Law says that the complement is distributive as long as you flip the sign in the middle.

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

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

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

Conditional Probability

Conditional probability is still probability, and also works as PDF!

Independent Events - **A** and **B** are independent if knowing one gives you no information about the other.

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

Also useful if X and Y are independent:

$$P(X = Y) = P(X = k)P(Y = k)$$

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.

Strategy: condition on the first step

Calvin and Hobbes play a match consisting of a series of games, where Calvin has probability p of winning each game independently. The first player to win two games more than his opponent wins the match.

To find the probability that Calvin wins a match , use LOTP, conditioning on the results of the first match. If the sequence is [CH] or [HC], then we're back at where we started. Then you solve for the probability you want.

To find the expected number of games played , consider each pair of games as a trial. Success: [HH] or [CC]. Failure: [HC] or [CH]. X = number of trials. $G = 2X$ = number of games. $X \sim FS(p)$ where $p = p^2 + q^2$.

Bayes Rule

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

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

When solving... do Bayes Rule first, and then do LOTP

Indicator Random Variables

Let I_j be the event that the jth letter in the sequence is an A.

Distribution $X_j \sim \text{Bern}(p)$ where $p = P(X_j = 1)$

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

Variance $Var(I_A) = p(1-p)$

Expectation

Expected Value - Values times probability for all possible values the r.v. can take on. If there are 3 buses carrying n students, then you should calculate expectation one by one (e.g. $\frac{1}{3}n_1 + \frac{1}{3}n_2 + \frac{1}{3}n_3$). Formally, for $X : \{x_1, x_2, x_3, \dots\}$:

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

Linearity For **any** random variables X and Y and any constants a, b, c, the following is true:

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

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

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

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

Independence If X and Y are independent, then

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

Using probability and expectation to prove existence

Possibility principle Let A before the event that a randomly chosen object in a collection has a certain property. If $P(A) > 0$, there there exists an object with the property

Good score principle Let X be the score of a randomly chosen object. If $E(X) \geq c$, then there is an object with a score of at least c

Continuous Random Variables

What's the prob that a CRV is in an interval?

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

Note that by the fundamental theorem of calculus,

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

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

$$P(4 < X < 16) = P(X < 16) - P(X < 4)$$
$$= \Phi\left(\frac{16-\mu}{\sigma^2}\right) - \Phi\left(\frac{4-\mu}{\sigma^2}\right)$$

Properties of valid CDF 1) F is increasing 2) F is right-continuous 3) $F(x) \to 1$ as $x \to \infty$, $F(x) \to 0$ as $x \to -\infty$

The PDF, $f(x)$, is the derivative of the CDF. It must integrate to 1 (because the probability that a CRV falls in the interval $[-\infty, \infty]$ is 1, and the PDF must always be nonnegative.

$$F'(x) = f(x)$$
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Law of the Unconscious Statistician (LotUS)

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? 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.

Universality of Uniform

For any X with CDF $F(x)$, $F(X) \sim U$.

Example Since you can have $\Phi(X) \sim unif(0, 1)$, then $E(\Phi(X)) = \frac{1}{2}$ using mean formula for uniform distribution.

Example Let $Z \sim N(0, 1)$. Create an Expo(1) r.v. X as a function (in terms of) Z. First, set CDF of Z equal to U. Then set Expo CDF equal to U and solve for X. Plug in U.

$$1 - e^{-\lambda X} = U \qquad \Phi(Z) = U$$
$$1 - U = e^{-X} \qquad X = -ln(1 - \Phi(Z))$$
$$ln(1 - U) = -X$$
$$X = -ln(1 - U)$$

Section 7 - Expo 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. The k^{th} moment of a random variable X is $E(X^k)$.

Skewness is third standardized moment

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E(X^3) - 3\mu E(X^2) + 2\mu^3}{\sigma^3}$$

Moment Generating Functions

MGF of X is the expected value and function of a dummy variable t for any r.v. X given by:

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

Discrete Example Let $X \sim Pois(\lambda)$:

$$M_X(t) = E(e^{tx}) = \sum_{x=0}^{\infty} \frac{e^{xt}e^{-\lambda}}{x!} = e^{\lambda(e^t-1)}$$

Continuous Example Suppose $X \sim Expo(\lambda)$. To find the moments, integrate for the MGF. Then plug in $t = 0$. Then differentiate for the next moment. Repeat for all the moments.

$$M_X(t) = E(e^{tx}) = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda - t} \text{ for } t < \lambda.$$

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!

$$E(X^k) = M_X^{(k)}(t = 0)$$

MGF of linear combination of X. If we have $Y = aX + c$, then

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

Summing Independent R.V.s by Multiplying MGFs. If X and Y are independent, then the sum of two random variables is the product of the MGFs of those two random variables.

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

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

Min and Max

One way is to differentiate and set equal to 0 to find the inflection points and also test the endpoints. Another way is with probabilities.

$$P(min(X, Y) \geq a) = P(X \geq a, Y \geq a)$$

Let i.i.d
$X_1, ..., X_n \sim Unif(0, 1)$ and $X_{(n)} = max(X_1, \ldots, X_n) \leq x)$. $X_{(n)}$ is less than x iff all of the X_j are less than x.

$$P(max(x_1, x_2, ...x_n) \geq w) = 1 - [P(x_1 \leq w)]^n = 1 - w^n$$

Joint PDFs and CDFs

Joint Distributions

Both the Joint PMF and Joint PDF must be non-negative and sum/integrate to 1. Defined as:

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

If independent, then join PMFs are just the marginal PMFs multiplied together!

Conditional Distributions

Hybrid Bayes' Rule is useful for finding posterior distributions of continuous r.v. conditioned on a discrete r.v. result. To solve a probability question that is binomial, but p has a continuous distribution, apply hybrid bayes rule, then use fundamental bridge and find the probability using expectation (Adam's law). Recall 2010 final asking $P(X_4 = 1)$ where $p \| (X = 3) \sim \text{beta}(a, b)$.

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

Marginal Distributions

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

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 Marginals
- 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:

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

Section 9 - Covariance

Variance

To solve, take out the constant first!

$$\text{Var}(cX) = c^2\text{Var}(X)$$

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

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

subsectionCovariance and Correlation (cont'd)

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, except in the case of Multivariate Normal, where uncorrelated *does* imply independence.

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

Covariance and Variance - Note that

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

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

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

In particular, If $X_1, X_2, \ldots, X_n$ are i.i.d. and all of them have the same covariance relationship, then

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = n\text{Var}(X_1) + 2\binom{n}{2}\text{Cov}(X_1, X_2)$$

Covariance and Linearity - For random variables W, X, Y, Z and constants b, c:

$$\text{Cov}(X + b, Y + c) = \text{Cov}(X, Y)$$

$$\text{Cov}(2X, 3Y) = 6\text{Cov}(X, Y)$$

$$\text{Cov}(W + X, Y + Z) = \text{Cov}(W, Y) + \text{Cov}(W, Z) + \text{Cov}(X, Y) + \text{Cov}(X,$$

Covariance and Invariance - Correlation, Covariance, and Variance are addition-invariant, which means that adding a constant to the term(s) does not change the value. Let b and c be constants.

$$\text{Var}(X + c) = \text{Var}(X)$$

$$\text{Cov}(X + b, Y + c) = \text{Cov}(X, Y)$$

$$\text{Corr}(X + b, Y + c) = \text{Corr}(X, Y)$$

In addition to addition-invariance, Correlation is *scale-invariant*, which means that multiplying the terms by any constant does not affect the value. Covariance and Variance are not scale-invariant.

$$\text{Corr}(2X, 3Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \text{Corr}(X, Y)$$

Tranformations (i.e. find the PDF)

If the question gives you two r.v., where you know the pdf of one r.v. and the other r.v. is a function, then the problem wants you do use transformation of variables. Solve for whatever variable you have on top. If it's dy/dx, solve for Y because it's on top. Create two r.v. for what is known and what is unknown. **The known r.v. should be a function of the unknown r.v.**. You can also find the pdf by differentiating the CDF.

One Variable Transformations If the function $g(X)$ is differentiable and every value of X gets mapped to a unique value of Y, then the following is true:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

Example $U \sim unif(0,1)$ and $Y = U^{1/a}$. Determine the PDF of Y by finding the jacobean, then plugging in the known PDF:

$$(Y)^a = (U^{1/a})^a$$
$$Y^a = U$$
$$\frac{dU}{dy} = |ay^{a-1}|$$

Two Variable Transformations Sometimes it's easier to do 1/J and then take the inverse.

$$f_{X,Y}(x,y) = f_{U,V}(u,v)\left|\left|\frac{\delta(u,v)}{\delta(x,y)}\right|\right| = f_{U,V}(u,v)\left|\left|\begin{array}{cc}\frac{\delta u}{\delta x} & \frac{\delta u}{\delta y}\\ \frac{\delta v}{\delta x} & \frac{\delta v}{\delta y}\end{array}\right|\right|$$

Remember to take the absolute value of the determinant matrix of partial derivatives. In a 2x2 matrix,

$$\left|\left|\begin{array}{cc} a & b\\ c & d\end{array}\right|\right| = |ad-bc|$$

Convolutions

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_x(x)f_y(t-x)dx$$

Example: Let $X, Y \sim i.i.dN(0,1)$. Treat t as a constant. Integrate as usual.

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2\pi}}e^{-(t-x)^2/2}dx$$

Conditional Expectation

Discrete Y	Continuous Y
$E(Y) = \sum_y yP(Y=y)$	$E(Y) = \int_{-\infty}^{\infty} yf_Y(y)dy$
$E(Y\|X=x) = \sum_y yP(Y=y\|X=x)$	$E(Y\|X=x) = \int_{-\infty}^{\infty} yf_{Y\|X}(y\|x)dy$
$E(Y\|A) = \sum_y yP(Y=y\|A)$	$E(Y\|A) = \int_{-\infty}^{\infty} yf(y\|A)dy$

Properties of Conditioning on Random Variables

1. $E(Y|X) = E(Y)$ if $X \perp\!\!\!\perp Y$
2. $E(h(X)|X) = h(X)$ (taking out what's known). $E(h(X)W|X) = h(X)E(W|X)$

Adam's Law

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

For any set of events that partition the sample space, $A_1, A_2, \ldots, A_n$ or just simply A, A^c, the following holds:

$$E(Y) = E(Y|A)P(A) + E(Y|A^c)P(A^c)$$
$$E(Y) = E(Y|A_1)P(A_1) + \cdots + E(Y|A_n)P(A_n)$$

Example Let $X \sim FS(p)$ and $p \sim Beta(a,b)$. $E(E(X|p)) = E(\frac{1}{p})$. Then use LOTUS and integrate. If integration looks painful, see if you can write the integral in terms of a PDF so that the integral equals 1.

Eve's Law

Eve's Law (aka Law of Total Variance)

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

Section 12 - MVN, LLN, CLT

Law of Large Numbers (LLN)

Let us have $X_1, X_2, X_3 \ldots$ be i.i.d.. We define

$$\bar{X}_n = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

The Law of Large Numbers states that as $n \longrightarrow \infty$, $\bar{X}_n \longrightarrow E(X)$

Example: Women's heights $\sim N(5.5, .025^2)$. A height of 5ft is 2σ away from the true population mean of 5.5ft. That means that the true population proportion of women shorter than 5 is 0.025. Based on the LLN, the sample proportion should be closer to the true population proportion of 2.5% more often when n is larger.

Central Limit Theorem (CLT)

Approximation using CLT

We use $\dot{\sim}$ to denote *is approximately distributed*. We can use the central limit theorem when we have a random variable, Y that is a sum of n i.i.d. random variables with n large. Let us say that $E(Y) = \mu_Y$ and $\text{Var}(Y) = \sigma_Y^2$. We have that:

$$Y \dot{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$$

When we use central limit theorem to estimate Y, we usually have $Y = X_1 + X_2 + \cdots + X_n$ or $Y = \bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$. Specifically, if we say that each of the iid X_i have mean μ_X and σ_X^2, then we have the following approximations.

$$X_1 + X_2 + \cdots + X_n \dot{\sim} \mathcal{N}(n\mu_X, n\sigma_X^2)$$
$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \dot{\sim} \mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$$

Asymptotic Distributions using CLT

We use $\xrightarrow{d}$ to denote *converges in distribution to* as $n \longrightarrow \infty$. These are the same results as the previous section, only letting $n \longrightarrow \infty$ and not letting our normal distribution have any n terms.

$$\frac{1}{\sigma\sqrt{n}}(X_1 + \cdots + X_n - n\mu_X) \xrightarrow{d} \mathcal{N}(0,1)$$

$$\frac{\bar{X}_n - \mu_X}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

Inequalities and Limit Theorems

Bounds on tail end probabilities

Markov

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

Chebychev If X and Y are independent, set $W = X - Y$ which has mean $E(W) = E(X) - E(Y) = 0$ to solve. Adding variables that sum to 0 is a good strategy to use.

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

Chernoff For any r.v. X and constants $a > 0$ and $t > 0$,

$$P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}$$

Bounds on Expectation

Remember that it's okay to replace expectation with equivalent expectation statements since expectation is just a number, but it's NOT okay to replace an r.v. inside of an expectation statement.

Cauchy-Schwarz $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$

Jensen

$$g \text{ convex} : E(g(X)) \geq g(E(X))$$
$$g \text{ concave} : E(g(X)) \leq g(E(X))$$
$$E|X| \geq |EX|$$
$$E\left(\frac{1}{X}\right) \geq \frac{1}{(EX)}, \text{ for positive r.v.s. X}$$
$$E(log(X)) \leq log(EX), \text{ for positive r.v.s. X}$$

Markov Chains

Definition

X_1 is a markov chain if its next state depends only on its current state. Formal Definition:

$P(X_{n+1} = a_{n+1}|X_1 = a_1, ..., X_n = a_n) = P(X_{n+1} = a_{n+1}|X_n = a_n)$

State Properties

- A state of MC is **recurrent** if, starting in that state, the MC will return to that state *eventually* with $p(\text{return eventually}) = 1$. Otherwise it is **transient**
- If every state is recurrent, this it's a recurrent chain. If a single state is transient, then it's a transient chain.
- The **period** of a state is the **greatest common denominator** of all the lengths from that state to itself. Periods only exist if the chain is **irreducible**. In an irreducible chain, all states have the same period.

Chain Properties

- A chain is **irreducible** if you can get from anywhere to anywhere. An irreducible chain must have all of its states recurrent.
- A chain is **aperiodic** if it has period 1 and is **periodic** otherwise.

A chain is **reversible** with respect to $\vec{s}$ if $s_i q_{ij} = s_j q_{ji}$ for all i, j. Vector of probabilities are same running forwards in time or backwards in time. Examples include random walks on undirected networks, or any chain with $q_{ij} = q_{ji}$, where the Markov chain would be stationary with respect to $\vec{s} = (\frac{1}{M}, \frac{1}{M}, \ldots, \frac{1}{M})$.

Reversibiity Condition Implies Stationary Distribution

Transition Matrix

Element q_{ij} in square transition matrix Q is the probability that the chain goes from state i to state j, or more formally:

$$q_{ij} = P(X_{n+1} = j | X_n = i)$$

To find the probability that the chain goes from state i to state j in m steps, take the $(i,j)^{\text{th}}$ element of Q^m.

$$q_{ij}^{(m)} = P(X_{n+m} = j | X_n = i)$$

If X_0 is distributed according to row-vector PMF $\vec{p}$ (e.g. $p_j = P(X_0 = i_j)$), then the PMF of X_n is $\vec{p}Q^n$.

Stationary Distribution

Let us say that the vector $\vec{p} = (p_1, p_2, \ldots, p_M)$ is a possible and valid PMF of where the Markov Chain is at at a certain time. We will call this vector the stationary distribution, $\vec{s}$, if it satisfies $\vec{s}Q = \vec{s}$. If X_t has the stationary distribution, then all future $X_{t+1}, X_{t+2}, \ldots$ also has the stationary distribution. To solve for the stationary distribution, you can solve for $(Q' - I)(\vec{s})' = 0$. The stationary distribution is uniform if the columns of Q sum to 1.

$$(s \quad 1-s) \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

Random Walk on Undirected Network

If you have a certain number of nodes with edges between them, and a chain can pick any edge randomly and move to another node, then this is a random walk on an undirected network. The stationary distribution of this chain is proportional to the **degree sequence**. The **degree sequence** is the vector of the degrees of each node, defined as how many edges it has.

Distribution Properties

Geometric Story

X is the number of "failures" that we will achieve before we achieve our first success. Our successes have probability p

$$1^2 - 2pq = p^2 + q^2$$

$$1^2 = (p+q)^2 = p^2 + 2pq + q^2$$

First Success

X is the number of "failures" that we will achieve until our first success, including the success, or number of games needed in order for someone to win the game (including the win). Our successes have probability p. If $X \sim \text{Geom}(p)$, then $X + 1 \sim FS(p)$.

Negative hypergeometric

PMF is $\frac{1}{g+1}$ where g is the event you want

Poisson Process

Definition We have a Poisson Process if we have

1. Arrivals at various times with an average of λ per unit time.
2. The number of arrivals in a time interval of length t is $\text{Pois}(\lambda t)$
3. Number of arrivals in disjoint time intervals are independent.

Normal

ALWAYS DRAW A PICTURE! For X is distributed $\mathcal{N}(\mu, \sigma^2)$, we know the following:

Symmetrical $Y = |X|$ means values are twice as likely to occur because both positive and negative values are counted. Answer is 2 times the PDF of the original normal distribution. Also, $P(Z^2 > 1) = P(Z < -1) + P(Z > 1) = \Phi(-1) + \Phi(-1)$. If it's a standard normal, use the 68-95-99.7 Rule.

Transformable $X = \mu + \sigma Z$ where bZ means the variance becomes b^2 and adding b just shifts the mean by b.

Standard Normal Remember it's $\sqrt{Var(x)}$!!!

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

68-95-99 rule $P(\|X - \mu\| < \sigma) = .68$, $P(\|X - \mu\| < 2\sigma) = .95$, and $P(\|X - \mu\| < 3\sigma) = 99.7$

Beta Properties

Say this "Beta is the prior conjugate of the binomial," if using it. If you condition on something with a likelihood that is binomial, the new r.v. will still be distributed beta (but with different parameters)

$$X|p \sim \text{Bin}(n,p)$$
$$p \sim \text{Beta}(a,b)$$

Then after observing the value $X = x$, we get a posterior distribution $p|(X = x) \sim \text{Beta}(a + x, b + n - x)$

Gamma

$\Gamma(a+1) = a\Gamma(a)$ for all $a > 0$

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

Poisson Properties (Chicken and Egg Results)

We have $X \sim \text{Pois}(\lambda_1)$ and $Y \sim \text{Pois}(\lambda_2)$ and $X \perp\!\!\!\perp Y$.

1. $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
2. $X|(X + Y = k) \sim \text{Bin}\left(k, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$
3. If we have that $Z \sim \text{Pois}(\lambda)$, and we randomly and independently "accept" every item in Z with probability p, then the number of accepted items $Z_1 \sim \text{Pois}(\lambda p)$, and the number of rejected items $Z_2 \sim \text{Pois}(\lambda q)$, and $Z_1 \perp\!\!\!\perp Z_2$.

Bank and Post Office Result

Let us say that we have $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$, and that $X \perp\!\!\!\perp Y$. By Bank-Post Office result, we have that:

$$X + Y \sim \text{Gamma}(a+b, \lambda)$$
$$\frac{X}{X+Y} \sim \text{Beta}(a,b)$$
$$X + Y \perp\!\!\!\perp \frac{X}{X+Y}$$

Memoryless Property

Geometric: If you're waiting for the first Heads in a sequence of fair coin tosses, the result of the previous tosses has no impact on the tosses we'll need. Expo: after you've waited s minutes, the probability you'll have to waiter another t minutes is exactly the same as when you'd just started waiting.

$$P(X \geq s+1 | X \geq s) = P(X \geq t)$$

How they all relate

1. $\text{Bin}(n,p) \to \text{Pois}(\lambda)$ as $n \to \infty$, $p \to 0$, $np = \lambda$.
2. sum of n i.i.d. expo with rate λ is gamma(n,λ).
3. sum of hgeom is NOT hgeom
4. $\text{NBin}(1,p) \sim \text{Geom}(p)$
5. $\text{Beta}(1,1) \sim \text{Unif}(0,1)$
6. $\text{Gamma}(1,\lambda) \sim \text{Expo}(\lambda)$
7. $\chi^2_n \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$

Convolutions of Random Variables

A convolution of n random variables is simply their sum.

1. $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2)$, $X \perp\!\!\!\perp Y \longrightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
2. $X \sim \text{Bin}(n_1, p)$, $Y \sim \text{Bin}(n_2, p)$, $X \perp\!\!\!\perp Y \longrightarrow X + Y \sim \text{Bin}(n_1 + n_2, p)$
3. $X \sim \text{Gamma}(n_1, \lambda)$, $Y \sim \text{Gamma}(n_2, \lambda)$, $X \perp\!\!\!\perp Y \longrightarrow X + Y \sim \text{Gamma}(n_1 + n_2, \lambda)$
4. $X \sim \text{NBin}(r_1, p)$, $Y \sim \text{NBin}(r_2, p)$, $X \perp\!\!\!\perp Y \longrightarrow X + Y \sim \text{NBin}(r_1 + r_2, p)$
5. All of the above are approximately normal when λ, n, r are large by the Central Limit Theorem.
6. $Z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Z_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $Z_1 \perp\!\!\!\perp Z_2 \longrightarrow Z_1 + Z_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

CDF/MGFs

Another method to find the CDF of a discrete r.v. to realize that $F(X) = 1 - P(X = 0)$

	CDF	MGF
Binomial	approx w/ normal for large n	$(q + pe^t)^n$
Geom	$1 - q^s$	$\frac{p}{1-qe^t}, qe^t < 1$
Unif	$\frac{x-a}{b-a}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
Expo	$1 - e^{-\lambda x}$	$\frac{\lambda}{\lambda - t}, t < \lambda$

Binomial, Bernoulli, Neg. Binomial, Geometric, Hypergeometric

DWR = Draw w/ replacement, DWoR = Draw w/o replacement

	DWR	DWoR
Fixed no. of trials (n)	Binom/Bern (Bern if $n = 1$)	HGeom
Draw until k successes	NBin/Geom (Geom if $k = 1$)	NHGeom

Continuous Distributions

Uniform Let us say that U is distributed Unif(a, b). We know the following:

Properties of the Uniform For a uniform distribution, the probability of an draw from any interval on the uniform is proportion to the length of the uniform. The PDF of a Uniform is just a constant, so when you integrate over the PDF, you will get an area proportional to the length of the interval.

Example William throws darts really badly, so his darts are uniform over the whole room because they're equally likely to appear anywhere. William's darts have a uniform distribution on the surface of the room. The uniform is the only distribution where the probably of hitting in any specific region is proportion to the area/length/volume of that region, and where the density of occurrence in any one specific spot is constant throughout the whole support.

PDF and CDF (top is Unif(0, 1), bottom is Unif(a, b))

$$f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases} \quad F(x) = \begin{cases} 0 & x < 0 \\ x & x \in [0,1] \\ 1 & x > 1 \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases} \quad F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$$

Normal Let us say that X is distributed $\mathcal{N}(\mu, \sigma^2)$. We know the following:

Central Limit Theorem The Normal distribution is ubiquitous because of the central limit theorem, which states that averages of independent identically-distributed variables will approach a normal distribution regardless of the initial distribution.

Transformable Every time we stretch or scale the normal distribution, we change it to another normal distribution. If we add c to a normally distributed random variable, then its mean increases additively by c. If we multiply a normally distributed random variable by c, then its variance increases multiplicatively by c^2. Note that for every normally distributed random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, we can transform it to the standard $\mathcal{N}(0, 1)$ by the following transformation:

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Example Heights are normal. Measurement error is normal. By the central limit theorem, the sampling average from a population is also normal.

Standard Normal - The Standard Normal, denoted Z, is $Z \sim \mathcal{N}(0, 1)$

PDF

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

CDF - It's too difficult to write this one out, so we express it as the function $\Phi(x)$

Exponential Distribution Let us say that X is distributed Expo(λ). We know the following:

Story You're sitting on an open meadow right before the break of dawn, wishing that airplanes in the night sky were shooting stars, because you could really use a wish right now. You know that shooting stars come on average every 15 minutes, but it's never true that a shooting star is ever ''due" to come because you've waited so long. Your waiting time is memorylessness, which means that the time until the next shooting star comes does not depend on how long you've waited already.

Example The waiting time until the next shooting star is distributed Expo(4). The 4 here is λ, or the rate parameter, or how many shooting stars we expect to see in a unit of time. The expected time until the next shooting star is $\frac{1}{\lambda}$, or $\frac{1}{4}$ of an hour. You can expect to wait 15 minutes until the next shooting star.

Expos are rescaled Expos

$$Y \sim \text{Expo}(\lambda) \to X = \lambda Y \sim \text{Expo}(1)$$

PDF and CDF The PDF and CDF of a Exponential is:

$$f(x) = \lambda e^{-\lambda x}, x \in [0, \infty)$$

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}, x \in [0, \infty)$$

Memorylessness The Exponential Distribution is the sole continuous memoryless distribution. This means that it's always "as good as new", which means that the probability of it failing in the next infinitesimal time period is the same as any infinitesimal time period. This means that for an exponentially distributed X and any real numbers t and s,

$$P(X > s + t | X > s) = P(X > t)$$

Given that you've waited already at least s minutes, the probability of having to wait an additional t minutes is the same as the probability that you have to wait more than t minutes to begin with. Here's another formulation.

$$X - a | X > a \sim \text{Expo}(\lambda)$$

Example - If waiting for the bus is distributed exponentially with $\lambda = 6$, no matter how long you've waited so far, the expected additional waiting time until the bus arrives is always $\frac{1}{6}$, or 10 minutes. The distribution of time from now to the arrival is always the same, no matter how long you've waited.

Gamma Distribution Let us say that X is distributed Gamma(a, λ). We know the following:

Story You sit waiting for shooting stars, and you know that the waiting time for a star is distributed Expo(λ). You want to see "a" shooting stars before you go home. X is the total waiting time for the ath shooting star.

Example You are at a bank, and there are 3 people ahead of you. The serving time for each person is distributed Exponentially with mean of 2 time units. The distribution of your waiting time until you begin service is Gamma(3, $\frac{1}{2}$)

PDF The PDF of a Gamma is:

$$f(x) = \frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} \frac{1}{x}, \quad x \in [0, \infty)$$

Properties and Representations

$$E(X) = \frac{a}{\lambda}, Var(X) = \frac{a}{\lambda^2}$$

$$X \sim G(a, \lambda), Y \sim G(b, \lambda), X \perp\!\!\!\perp Y \to X+Y \sim G(a+b, \lambda), \frac{X}{X+Y} \perp\!\!\!\perp X+$$

$$X \sim \text{Gamma}(a, \lambda) \to X = X_1 + X_2 + ... + X_a \text{ for } X_i \text{ i.i.d. Expo}(\lambda)$$

$$\text{Gamma}(1, \lambda) \sim \text{Expo}(\lambda)$$

χ^2 Distribution Let us say that X is distributed χ^2_n. We know the following:

Story A Chi-Squared(n) is a sum of n independent squared normals.

Example The sum of squared errors are distributed χ^2_n

PDF The PDF of a χ^2_1 is:

$$f(w) = \frac{1}{\sqrt{2\pi w}} e^{-w/2}, w \in [0, \infty)$$

Properties and Representations

$$E(\chi^2_n) = n, Var(X) = 2n$$

$$\chi^2_n \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$\chi^2_n = Z_1^2 + Z_2^2 + \cdots + Z_n^2, Z \sim^{i.i.d.} \mathcal{N}(0, 1)$$

Discrete Distributions

Bernoulli The Bernoulli distribution is the simplest case of the Binomial distribution, where we only have one trial, or $n = 1$. Let us say that X is distributed Bern(p). We know the following:

Story. X "succeeds" (is 1) with probability p, and X "fails" (is 0) with probability $1 - p$.

Example. A fair coin flip is distributed Bern($\frac{1}{2}$).

PMF. The probability mass function of a Bernoulli is:

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

or simply

$$P(X = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

Binomial Let us say that X is distributed Bin(n, p). We know the following:

Story X is the number of "successes" that we will achieve in n independent trials, where each trial can be either a success or a failure, each with the same probability p of success. We can also say that X is a sum of multiple independent $Bern(p)$ random variables. Let $X \sim \text{Bin}(n,p)$ and $X_j \sim \text{Bern}(p)$, where all of the Bernoullis are independent. We can express the following:

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

Example If Jeremy Lin makes 10 free throws and each one independently has a $\frac{3}{4}$ chance of getting in, then the number of free throws he makes is distributed $\text{Bin}(10, \frac{3}{4})$, or, letting X be the number of free throws that he makes, X is a Binomial Random Variable distributed $\text{Bin}(10, \frac{3}{4})$.

PMF The probability mass function of a Binomial is:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Binomial Coefficient $\binom{n}{k}$ is a function of n and k and is read n *choose* k, and means out of n possible indistinguishable objects, how many ways can I possibly choose k of them? The formula for the binomial coefficient is:

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

Geometric Let us say that X is distributed $\text{Geom}(p)$. We know the following:

Story X is the number of "failures" that we will achieve before we achieve our first success. Our successes have probability p.

Example If each pokeball we throw has a $\frac{1}{10}$ probability to catch Mew, the number of failed pokeballs will be distributed $\text{Geom}(\frac{1}{10})$.

PMF With $q = 1-p$, the probability mass function of a Geometric is:

$$P(X = k) = q^k p$$

Negative Binomial Let us say that X is distributed $\text{NBin}(r,p)$. We know the following:

Story X is the number of "failures" that we will achieve before we achieve our rth success. Our successes have probability p.

Example Thundershock has 60% accuracy and can faint a wild Raticate in 3 hits. The number of misses before Pikachu faints Raticate with Thundershock is distributed $\text{NBin}(3, .6)$.

PMF With $q = 1-p$, the probability mass function of a Negative Binomial is:

$$P(X = n) = \binom{n+r-1}{r-1} p^r q^n$$

Hypergeometric Let us say that X is distributed $\text{HGeom}(w,b,n)$. We know the following:

Story In a population of b undesired objects and w desired objects, X is the number of "successes" we will have in a draw of n objects, without replacement.

Example 1) Let's say that we have only b Weedles (failure) and w Pikachus (success) in Viridian Forest. We encounter n Pokemon in the forest, and X is the number of Pikachus in our encounters. 2) The number of aces that you draw in 5 cards (without replacement). 3) You have w white balls and b black balls, and you draw b balls. You will draw X white balls. 4) Elk Problem - You have N elk, you capture n of them, tag them, and release them. Then you recollect a new sample of size m. How many tagged elk are now in the new sample?

PMF The probability mass function of a Hypergeometric:

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

Poisson Let us say that X is distributed $\text{Pois}(\lambda)$. We know the following:

Story There are rare events (low probability events) that occur many different ways (high possibilities of occurences) at an average rate of λ occurrences per unit space or time. The number of events that occur in that unit of space or time is X.

Example A certain busy intersection has an average of 2 accidents per month. Since an accident is a low probability event that can happen many different ways, the number of accidents in a month at that intersection is distributed Pois(2). The number of accidents that happen in two months at that intersection is distributed Pois(4)

PMF The PMF of a Poisson is:

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

Multivariate Distributions

Multinomial Let us say that the vector $\vec{\mathbf{X}} = (X_1, X_2, X_3, \ldots, X_k) \sim \text{Mult}_k(n, \vec{p})$ where $\vec{p} = (p_1, p_2, \ldots, p_k)$.

Story - We have n items, and then can fall into any one of the k buckets independently with the probabilities $\vec{p} = (p_1, p_2, \ldots, p_k)$.

Example - Let us assume that every year, 100 students in the Harry Potter Universe are randomly and independently sorted into one of four houses with equal probability. The number of people in each of the houses is distributed $\text{Mult}_4(100, \vec{p})$, where $\vec{p} = (.25, .25, .25, .25)$. Note that $X_1 + X_2 + \cdots + X_4 = 100$, and they are dependent.

Multinomial Coefficient The number of permutations of n objects where you have $n_1, n_2, n_3 \ldots, n_k$ of each of the different variants is the **multinomial coefficient**.

$$\binom{n}{n_1 n_2 \ldots n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}$$

Joint PMF - For $n = n_1 + n_2 + \cdots + n_k$

$$P(\vec{X} = \vec{n}) = \binom{n}{n_1 n_2 \ldots n_k} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

Lumping - If you lump together multiple categories in a multinomial, then it is still multinomial. A multinomial with two dimensions (success, failure) is a binomial distribution.

Variances and Covariances - For $(X_1, X_2, \ldots, X_k) \sim \text{Mult}_k(n, (p_1, p_2, \ldots, p_k))$, we have that marginally $X_i \sim \text{Bin}(n, p_i)$ and hence $\text{Var}(X_i) = np_i(1-p_i)$. Also, for $i \neq j$, $\text{Cov}(X_i, X_j) = -np_ip_j$, which is a result from class.

Marginal PMF and Lumping

$$X_i \sim \text{Bin}(n, p_i)$$

$$X_i + X_j \sim \text{Bin}(n, p_i + p_j)$$

$$X_1, X_2, X_3 \sim \text{Mult}_3(n, (p_1, p_2, p_3)) \to X_1, X_2 + X_3 \sim \text{Mult}_2(n, (p_1, p_2 + p_3))$$

$$X_1, \ldots, X_{k-1} | X_k = n_k \sim \text{Mult}_{k-1}\left(n - n_k, \left(\frac{p_1}{1-p_k}, \ldots, \frac{p_{k-1}}{1-p_k}\right)\right)$$

Multivariate Uniform See the univariate uniform for stories and examples. For multivariate uniforms, all you need to know is that probability is proportional to volume. More formally, probability is the volume of the region of interest divided by the total volume of the support. Every point in the support has equal density of value $\frac{1}{\text{Total Area}}$.

Multivariate Normal (MVN) A vector $\vec{X} = (X_1, X_2, X_3, \ldots, X_k)$ is declared Multivariate Normal if any linear combination is normally distributed (e.g. $t_1X_1 + t_2X_2 + \cdots + t_kX_k$ is Normal for any constants $t_1, t_2, \ldots, t_k$). The parameters of the Multivariate normal are the mean vector $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ and the covariance matrix where the $(i,j)^{\text{th}}$ entry is $\text{Cov}(X_i, X_j)$. For any MVN distribution: 1) Any sub-vector is also MVN. 2) If any two elements of a multivariate normal distribution are uncorrelated, then they are independent. Note that 2) does not apply to most random variables.

Compiled by William Chen '14. Please email comments, suggestions, and errors to williamzc@gmail.com

Distribution	PDF and Support	EV	Variance	MGF
Bernoulli Bern(p)	$P(X=1)=p$ $P(X=0)=q$	p	pq	$q+pe^t$
Binomial Bin(n,p)	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ $k\in\{0,1,2,\dots n\}$	np	npq	$(q+pe^t)^n$
Geometric Geom(p)	$P(X=k)=q^kp$ $k\in\{0,\ 1,\ 2,\ \dots\}$	q/p	q/p^2	$\frac{p}{1-qe^t}, qe^t<1$
Negative Binom. NBin(r,p)	$P(X=n)=\binom{r+n-1}{r-1}p^rq^n$ $n\in\{0,\ 1,\ 2,\ \dots\}$	rq/p	rq/p^2	$(\frac{p}{1-qe^t})^r, qe^t<1$
Hypergeometric HGeom(w,b,n)	$P(X=k)=\binom{w}{k}\binom{b}{n-k}/\binom{w+b}{n}$ $k\in\{0,1,2,\dots,n\}$	$\mu=\frac{nw}{b+w}$	$\frac{w+b-n}{w+b-1}n\frac{\mu}{n}(1-\frac{\mu}{n})$	–
Poisson Pois(λ)	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$ $k\in\{0,\ 1,\ 2,\ \dots\}$	λ	λ	$e^{\lambda(e^t-1)}$
Uniform Unif(a,b)	$f(x)=\frac{1}{b-a}$ $x\in(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Normal $\mathcal{N}(\mu,\sigma^2)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$ $x\in(-\infty,\infty)$	μ	σ^2	$e^{t\mu+\frac{\sigma^2t^2}{2}}$
Exponential Expo(λ)	$f(x)=\lambda e^{-\lambda x}$ $x\in(0,\infty)$	$1/\lambda$	$1/\lambda^2$	$\frac{\lambda}{\lambda-t}, t<\lambda$
Gamma Gamma(a,λ)	$f(x)=\frac{1}{\Gamma(a)}(\lambda x)^ae^{-\lambda x}\frac{1}{x}$ $x\in(0,\infty)$	a/λ	a/λ^2	$\left(\frac{\lambda}{\lambda-t}\right)^a, t<\lambda$
Beta Beta(a, b)	$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$ $x\in(0,1)$	$\mu=\frac{a}{a+b}$	$\frac{\mu(1-\mu)}{(a+b+1)}$	–
Chi-Squared χ^2_n	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$ $x\in(0,\infty)$	n	$2n$	$(1-2t)^{-n/2}, t<1/2$
Multivar Uniform A is support	$f(x)=\frac{1}{\lvert A\rvert}$ $x\in A$	–	–	–
Multinomial Mult$_k(n,\vec{p})$	$P(\vec{X}=\vec{n})=\binom{n}{n_1\dots n_k}p_1^{n_1}\dots p_k^{n_k}$ $n=n_1+n_2+\dots+n_k$	$n\vec{p}$	$\mathrm{Var}(X_i)=np_i(1-p_i)$ $\mathrm{Cov}(X_i,X_j)=-np_ip_j$	$\left(\sum_{i=1}^k p_ie^{t_i}\right)^n$

Cauchy-Schwarz	Markov	Chebychev	Jensen
$\lvert E(XY)\rvert\le\sqrt{E(X^2)E(Y^2)}$	$P(X\ge a)\le\frac{E\lvert X\rvert}{a}$	$P(\lvert X-\mu_X\rvert\ge a)\le\frac{\sigma_X^2}{a^2}$	g convex: $E(g(X))\ge g(E(X))$ g concave: $E(g(X))\le g(E(X))$