Stat 110 Strategic Practice 9, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1 Beta and Gamma Distributions

1. Let $B \sim \text{Beta}(a, b)$. Find the distribution of $1 - B$ in two ways: (a) using a change of variables and (b) using a story proof. Also explain why the result makes sense in terms of Beta being the conjugate prior for the Binomial.

2. Let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$ be independent, with a and b integers. Show that $X + Y \sim \text{Gamma}(a + b, \lambda)$ in three ways: (a) with a convolution integral; (b) with MGFs; (c) with a story proof.

3. Fred waits $X \sim \text{Gamma}(a, \lambda)$ minutes for the bus to work, and then waits $Y \sim \text{Gamma}(b, \lambda)$ for the bus going home, with X and Y independent. Is the ratio X/Y independent of the total wait time $X + Y$?

4. The F-test is a very widely-used statistical test based on the $F(m, n)$ distribution, which is the distribution of $\frac{X/m}{Y/n}$ with $X \sim \text{Gamma}(\frac{m}{2}, \frac{1}{2}), Y \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$. Find the distribution of $mV/(n + mV)$ for $V \sim F(m, n)$.

2 Order Statistics

1. Let $U_1, \dots, U_n$ be i.i.d. $\text{Unif}(0, 1)$. Find the PDF of the jth order statistic $U_{(j)}$ (including the normalizing constant), and its mean and variance.

2. Let X and Y be independent $\text{Expo}(\lambda)$ r.v.s and $M = \max(X, Y)$. Show that M has the same distribution as $X + \frac{1}{2}Y$, in two ways: (a) using calculus and (b) by remembering the memoryless property and other properties of the Exponential.

3. Let $X_1, X_2, \dots, X_n$ be i.i.d. r.v.s with CDF F and PDF f. Find the joint PDF of the order statistics $X_{(i)}$ and $X_{(j)}$ for $1 \le i < j \le n$, by drawing and thinking about a picture.

3 Conditional Expectation

1. You get to choose between two envelopes, each of which contains a check for some positive amount of money. Unlike in the two envelope paradox from class,

it is not given that one envelope contains twice as much money as the other envelope. Instead, assume that the two values were generated independently from some distribution on positive real numbers, with no information given about what that distribution is.

After picking an envelope, you can open it and see how much money is inside (call this value x), and then you have the option of switching. As no information has been given about the distribution, it may seem impossible to have better than a 50% chance of picking the better envelope. Intuitively, we may want to switch if x is "small" and not switch if x is "large," but how do we define "small" and "large" in the grand scheme of all possible distributions? [The last sentence was a rhetorical question.]

Consider the following strategy for deciding whether to switch. Generate a "threshold" $T \sim \text{Expo}(1)$, and switch envelopes if and only if the observed value x is less than the value of T. Show that this strategy succeeds in picking the envelope with more money with probability strictly greater than 1/2.

Hint: Let t be the value of T (generated by a random draw from the Expo(1) distribution). First explain why the strategy works very well if t happens to be in between the two envelope values, and does no harm in any case (i.e., there is no case in which the strategy succeeds with probability strictly less than 1/2).

2. Give a short, direct proof that a Geom(p) r.v. has mean q/p, for $q = 1 - p$.

3. The "Mass Cash" lottery randomly chooses 5 of the numbers from $1, 2, \ldots, 35$ each day (without repetitions within the choice of 5 numbers). Suppose that we want to know how long it will take until all numbers have been chosen. Let a_j be the average number of additional days needed if we are missing j numbers (so $a_0 = 0$ and a_{35} is the average number of days needed to collect all 35 numbers). Find a recursive formula for the a_j's.

4. You are given an amazing opportunity to bid on a mystery box containing a mystery prize! The value of the prize is completely unknown, except that it is worth at least nothing, and at most a million dollars. So the true value V of the prize is considered to be Uniform on [0,1] (measured in millions of dollars).

   You can choose to bid any amount b (in millions of dollars). You have the chance to get the prize for considerably less than it is worth, but you could also lose money if you bid too much. Specifically, if $b < \frac{2}{3}V$, then the bid is rejected and nothing is gained or lost. If $b \geq \frac{2}{3}V$, then the bid is accepted and your net payoff is $V - b$ (since you pay b to get a prize worth V). What is your optimal bid b (to maximize the expected payoff)?

Stat 110 Strategic Practice 9 Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1 Beta and Gamma Distributions

1. Let $B \sim \text{Beta}(a, b)$. Find the distribution of $1 - B$ in two ways: (a) using a change of variables and (b) using a story proof. Also explain why the result makes sense in terms of Beta being the conjugate prior for the Binomial.

   (a) Let $W = 1 - B$. The function $g(t) = 1 - t$ is strictly decreasing with absolute derivative $|-1| = 1$, so the PDF of W is

   $$f_W(w) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(1-w)^{a-1}w^{b-1},$$

   for $0 < w < 1$, which shows that $W \sim \text{Beta}(b, a)$.

   (b) Using the bank-post office story, we can represent $B = \frac{X}{X+Y}$ with $X \sim \text{Gamma}(a, 1)$ and $Y \sim \text{Gamma}(b, 1)$ independent. Then $1 - B = \frac{Y}{X+Y} \sim \text{Beta}(b, a)$ by the same story.

   This result makes sense intuitively since if we use $\text{Beta}(a, b)$ as the prior distribution for the probability p of success in a Binomial problem, interpreting a as the number of prior successes and b as the number of prior failures, then $1 - p$ is the probability of failure and, interchanging the roles of "success" and "failure," it makes sense to have $1 - p \sim \text{Beta}(b, a)$.

2. Let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$ be independent, with a and b integers. Show that $X + Y \sim \text{Gamma}(a + b, \lambda)$ in three ways:

   (a) with a convolution integral;

   The convolution integral is

   $$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t-x)dx = \int_0^t \frac{1}{\Gamma(a)}\frac{1}{\Gamma(b)}(\lambda x)^a(\lambda(t-x))^b e^{-\lambda x}e^{-\lambda(t-x)}\frac{1}{x}\frac{1}{t-x}dx,$$

   where we integrate from 0 to t since we need $x > 0$ and $t - x > 0$. This is

   $$\lambda^{a+b}\frac{e^{-\lambda t}}{\Gamma(a)\Gamma(b)}\int_0^t x^{a-1}(t-x)^{b-1}dx = \lambda^{a+b}\frac{e^{-\lambda t}}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}t^{a+b-1} = \frac{1}{\Gamma(a+b)}(\lambda t)^{a+b}e^{-\lambda t}\frac{1}{t},$$

using a Beta integral (after letting $u = x/t$ so that we can integrate from 0 to 1 rather than 0 to t). Thus, $X + Y \sim \text{Gamma}(a + b, \lambda)$.

(b) with MGFs;

The MGF of $X+Y$ is $M_X(t)M_Y(t) = \frac{\lambda^a}{(\lambda-t)^a}\frac{\lambda^b}{(\lambda-t)^b} = \frac{\lambda^{a+b}}{(\lambda-t)^{a+b}} = M_{X+Y}(t)$, which again shows that $X + Y \sim \text{Gamma}(a + b, \lambda)$.

(c) with a story proof.

Interpret X as the time of the ath arrival in a Poisson process with rate λ, and Y as the time needed for b more arrivals to occur (which is independent of X since the times between arrivals are independent $\text{Expo}(\lambda)$ r.v.s). Then $X + Y$ is the time of the $(a + b)$th arrival, so $X + Y \sim \text{Gamma}(a + b, \lambda)$.

3. Fred waits $X \sim \text{Gamma}(a, \lambda)$ minutes for the bus to work, and then waits $Y \sim \text{Gamma}(b, \lambda)$ for the bus going home, with X and Y independent. Is the ratio X/Y independent of the total wait time $X + Y$?

   As shown in the bank-post office story, $W = \frac{X}{X+Y}$ is independent of $X + Y$. So any function of W is independent of any function of $X + Y$. And we have that X/Y is a function of W, since

$$\frac{X}{Y} = \frac{\frac{X}{X+Y}}{\frac{Y}{X+Y}} = \frac{W}{1-W},$$

   so X/Y is independent of $X + Y$.

4. The F-test is a very widely-used statistical test based on the $F(m, n)$ distribution, which is the distribution of $\frac{X/m}{Y/n}$ with $X \sim \text{Gamma}(\frac{m}{2}, \frac{1}{2}), Y \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$. Find the distribution of $mV/(n + mV)$ for $V \sim F(m, n)$.

   Let $X \sim \text{Gamma}(\frac{m}{2}, \frac{1}{2}), Y \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$, and $V = \frac{n}{m}\frac{X}{Y}$. Then

$$mV/(n + mV) = \frac{nX/Y}{n + nX/Y} = \frac{X}{X + Y} \sim \text{Beta}\left(\frac{m}{2}, \frac{n}{2}\right).$$

2 Order Statistics

1. Let $U_1, \ldots, U_n$ be i.i.d. Unif$(0,1)$. Find the PDF of the jth order statistic $U_{(j)}$ (including the normalizing constant), and its mean and variance.

   In class it was shown that $U_{(j)} \sim \text{Beta}(j, n-j+1)$, so the PDF of $U_{(j)}$ is

   $$f(t) = \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} t^{j-1}(1-t)^{n-j} = \frac{n!}{(j-1)!(n-j)!} t^{j-1}(1-t)^{n-j},$$

   for $0 < t < 1$ (and 0 otherwise). The mean of a r.v. $T \sim \text{Beta}(a,b)$ was shown in class to be $\frac{a}{a+b}$. To get the variance, first compute the second moment of T:

   $$\int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^2 t^{a-1}(1-t)^{b-1} dt = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 t^{a+1}(1-t)^{b-1} dt,$$

   where we're integrating the Beta$(a+2, b)$ PDF (after normalizing), so

   $$E(T^2) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} = \frac{(a+1)a}{(a+b+1)(a+b)},$$

   using the fact that $\Gamma(x+1) = x\Gamma(x)$. So

   $$\text{Var}(T) = \frac{(a+1)a}{(a+b+1)(a+b)} - \left(\frac{a}{a+b}\right)^2 = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\mu(1-\mu)}{a+b+1},$$

   where $\mu = a/(a+b)$. Therefore, $U_{(j)}$ has mean $\mu = \frac{j}{n+1}$ and variance $\frac{\mu(1-\mu)}{n+2} = \frac{j(n-j+1)}{(n+1)^2(n+2)}$.

2. Let X and Y be independent Expo(λ) r.v.s and $M = \max(X, Y)$. Show that M has the same distribution as $X + \frac{1}{2}Y$, in two ways: (a) using calculus and (b) by remembering the memoryless property and other properties of the Exponential.

   (a) The CDF of M is

   $$F_M(x) = P(M \leq x) = P(X \leq x, Y \leq x) = (1 - e^{-\lambda x})^2,$$

   and the CDF of $X + \frac{1}{2}Y$ is

   $$F_{X+\frac{1}{2}Y}(x) = P(X + \frac{1}{2}Y \leq x) = \iint\limits_{s+\frac{1}{2}t \leq x} \lambda^2 e^{-\lambda s - \lambda t} ds dt$$

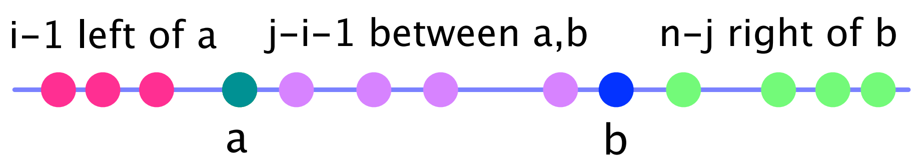
$$= \int_0^{2x} \lambda e^{-\lambda t} dt \int_0^{x-\frac{1}{2}t} \lambda e^{-\lambda s} ds$$

$$= \int_0^{2x} (1 - e^{-\lambda x - \frac{1}{2}t}) \lambda e^{-\lambda t} dt = (1 - e^{-\lambda x})^2.$$

Thus, M and $X + \frac{1}{2}Y$ have the same CDF.

(b) This is similar to the "3 students working on a pset" problem, with 2 students instead of 3. Let $L = \min(X, Y)$ and write $M = L + (M - L)$. In that story (with 2 students instead of 3), $L \sim \text{Expo}(2\lambda)$ is the time it takes for the first student to finish the pset and then by the memoryless property, the additional time until the second student finishes the pset is $M - L \sim \text{Expo}(\lambda)$, independent of L. Since $\frac{1}{2}Y \sim \text{Expo}(2\lambda)$ is independent of $X \sim \text{Expo}(\lambda)$, $M = L + (M - L)$ has the same distribution as $\frac{1}{2}Y + X$.

3. Let $X_1, X_2, \ldots, X_n$ be i.i.d. r.v.s with CDF F and PDF f. Find the joint PDF of the order statistics $X_{(i)}$ and $X_{(j)}$ for $1 \le i < j \le n$, by drawing and thinking about a picture.



To have $X_{(i)}$ be in a tiny interval around a and $X_{(j)}$ be in a tiny interval around b, where $a < b$, we need to have 1 of the X_k's be almost exactly at a, another be almost exactly at b, $i - 1$ of them should be to the left of a, $n - j$ should be to the right of b, and the remaining $j - i - 1$ should be between a and b, as shown in the picture. This gives that the PDF is

$$f_{(i),(j)}(a, b) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(a)^{i-1} f(a) (F(b) - F(a))^{j-i-1} f(b) (1 - F(b))^{n-j},$$

for $a < b$. The coefficient in front counts the number of ways to put the X_k's into the 5 categories "left of a," "at a," "between a and b," "at b," "right of b" with the desired number in each category (which is the same idea used to find the coefficient in front of the Multinomial PMF). Equivalently, we could write the coefficient as $n(n-1)\binom{n-2}{i-1}\binom{n-i-1}{j-i-1}$, since there are n choices for which X_k is at a, then $n - 1$ choices for which is at b, etc.

3 Conditional Expectation

1. You get to choose between two envelopes, each of which contains a check for some positive amount of money. Unlike in the two envelope paradox from class, it is not given that one envelope contains twice as much money as the other envelope. Instead, assume that the two values were generated independently from some distribution on positive real numbers, with no information given about what that distribution is.

   After picking an envelope, you can open it and see how much money is inside (call this value x), and then you have the option of switching. As no information has been given about the distribution, it may seem impossible to have better than a 50% chance of picking the better envelope. Intuitively, we may want to switch if x is "small" and not switch if x is "large," but how do we define "small" and "large" in the grand scheme of all possible distributions? [The last sentence was a rhetorical question.]

   Consider the following strategy for deciding whether to switch. Generate a "threshold" $T \sim \text{Expo}(1)$, and switch envelopes if and only if the observed value x is less than the value of T. Show that this strategy succeeds in picking the envelope with more money with probability strictly greater than 1/2.

   Hint: Let t be the value of T (generated by a random draw from the Expo(1) distribution). First explain why the strategy works very well if t happens to be in between the two envelope values, and does no harm in any case (i.e., there is no case in which the strategy succeeds with probability strictly less than 1/2).

   Let a be the smaller value of the two envelopes and b be the larger value (assume $a < b$ since in the case $a = b$ it makes no difference which envelope is chosen!). Let G be the event that the strategy succeeds and A be the event that we pick the envelope with a initially. Then $P(G|A) = P(T > a) = 1 - (1 - e^{-a}) = e^{-a}$, and $P(G|A^c) = P(T \le b) = 1 - e^{-b}$. Thus, the probability that the strategy succeeds is
   $$\frac{1}{2}e^{-a} + \frac{1}{2}(1 - e^{-b}) = \frac{1}{2} + \frac{1}{2}(e^{-a} - e^{-b}) > \frac{1}{2},$$
   because $e^{-a} - e^{-b} > 0$.

2. Give a short, direct proof that a Geom(p) r.v. has mean q/p, for $q = 1 - p$.

   Let $X \sim \text{Geom}(p)$, interpreted as the number of failures before the first success in independent Bern(p) trials. Let A be the event that the first trial is a success.

Then

$$E(X) = E(X|A)P(A) + E(X|A^c)P(A^c) = 0p + (1 + E(X))q$$

yields $E(X) = q/p$.

3. The "Mass Cash" lottery randomly chooses 5 of the numbers from $1, 2, \ldots, 35$ each day (without repetitions within the choice of 5 numbers). Suppose that we want to know how long it will take until all numbers have been chosen. Let a_j be the average number of additional days needed if we are missing j numbers (so $a_0 = 0$ and a_{35} is the average number of days needed to collect all 35 numbers). Find a recursive formula for the a_j's.

   Suppose we are missing j numbers (with $0 \leq j \leq 35$), and let T_j be the additional number of days needed to complete the collection. Condition on how many "new" numbers appear the next day; call this N. This gives

$$E(T_j) = \sum_{n=0}^{5} E(T_j|N = n)P(N = n).$$

   Note that N is Hypergeometric (imagine tagging the numbers that we don't already have in our collection)! Letting $a_k = 0$ for $k < 0$, we have

$$a_j = 1 + \sum_{n=0}^{5} \frac{a_{j-n}\binom{j}{n}\binom{35-j}{5-n}}{\binom{35}{5}}.$$

4. You are given an amazing opportunity to bid on a mystery box containing a mystery prize! The value of the prize is completely unknown, except that it is worth at least nothing, and at most a million dollars. So the true value V of the prize is considered to be Uniform on [0,1] (measured in millions of dollars).

   You can choose to bid any amount b (in millions of dollars). You have the chance to get the prize for considerably less than it is worth, but you could also lose money if you bid too much. Specifically, if $b < \frac{2}{3}V$, then the bid is rejected and nothing is gained or lost. If $b \geq \frac{2}{3}V$, then the bid is accepted and your net payoff is $V - b$ (since you pay b to get a prize worth V). What is your optimal bid b (to maximize the expected payoff)?

   We choose a bid $b \geq 0$, which cannot be defined in terms of the unknown V. The expected payoff can be found by conditioning on whether the bid is accepted. The term where the bid is rejected is 0, so the expected payoff is

$$E(V - b | b \geq \frac{2}{3}V) P(b \geq \frac{2}{3}V) = \left(E(V | V \leq \frac{3}{2}b) - b\right) P(V \leq \frac{3}{2}b).$$

For $b \geq 2/3$, the bid is definitely accepted but we lose money on average, so assume $b < 2/3$. Then

$$\left(E(V | V \leq \frac{3}{2}b) - b\right) P(V \leq \frac{3}{2}b) = (\frac{3}{4}b - b)\frac{3}{2}b = -\frac{3}{8}b^2,$$

since given that $V \leq \frac{3}{2}b$, the conditional distribution of V is Uniform on $[0, \frac{3}{2}b]$.

The above expression is negative except at $b = 0$, so the optimal bid is 0: one should not play this game! What's the moral of this story? First, investing in an asset without any information about its value is a bad idea. Second, *condition on all the information*. It is crucial in the above calculation to use $E(V | V \leq \frac{3}{2}b)$ rather than $E(V) = 1/2$; knowing that the bid was accepted gives information about how much the mystery prize is worth!

Stat 110 Homework 9, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. Let X and Y be independent, positive r.v.s. with finite expected values.

(a) Give an example where $E(\frac{X}{X+Y}) \neq \frac{E(X)}{E(X+Y)}$, computing both sides exactly.
Hint: start by thinking about the simplest examples you can think of!

(b) If X and Y are i.i.d., then is it necessarily true that $E(\frac{X}{X+Y}) = \frac{E(X)}{E(X+Y)}$?

(c) Now let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$. Show *without using calculus* that $E(\frac{X^c}{(X+Y)^c}) = \frac{E(X^c)}{E((X+Y)^c)}$ for every real $c > 0$.

2. The *Gumbel distribution* is the distribution of $-\log X$ with $X \sim \text{Expo}(1)$.

(a) Find the CDF of the Gumbel distribution.

(b) Let $X_1, X_2, \ldots$ be i.i.d. Expo(1) and let $M_n = \max(X_1, \ldots, X_n)$. Show that $M_n - \log n$ converges in distribution to the Gumbel distribution, i.e., as $n \to \infty$ the CDF of $M_n - \log n$ converges to the Gumbel CDF.

3. Consider independent Bernoulli trials with probability p of success for each. Let X be the number of failures incurred before getting a total of r successes.

(a) Determine what happens to the distribution of $\frac{p}{1-p}X$ as $p \to 0$, using MGFs; what is the PDF of the limiting distribution, and its name and parameters if it is one we have studied?
Hint: start by finding the Geom(p) MGF. Then find the MGF of $\frac{p}{1-p}X$, and use the fact that if the MGFs of r.v.s Y_n converge to the MGF of a r.v. Y, then the CDFs of the Y_n converge to the CDF of Y.

(b) Explain intuitively why the result of (a) makes sense.

4. (a) If X and Y are i.i.d. continuous r.v.s with CDF $F(x)$ and PDF $f(x)$, then $M = \max(X, Y)$ has PDF $2F(x)f(x)$. Now let X and Y be discrete and i.i.d., with CDF $F(x)$ and PMF $f(x)$. Explain in words why the PMF of M is *not* $2F(x)f(x)$.

(b) Let X and Y be independent Bernoulli(1/2) r.v.s, and let $M = \max(X, Y)$, $L = \min(X, Y)$. Find the joint PMF of M and L, i.e., $P(M = a, L = b)$, and the marginal PMFs of M and L.

5. Let $X \sim \text{Bin}(n, p)$ and $B \sim \text{Beta}(j, n - j + 1)$, where n is a positive integer and j is a positive integer with $j \leq n$. Show using a story about order statistics (without using calculus) that

$$P(X \geq j) = P(B \leq p).$$

This shows that the CDF of the continuous r.v. B is closely related to the CDF of the discrete r.v. X, and is another connection between the Beta and Binomial.

6. A coin with probability p of Heads is flipped repeatedly. For Parts (a) and (b), suppose that p is a known constant, with $0 < p < 1$.

(a) What is the expected number of flips until the pattern HT is observed?

(b) What is the expected number of flips until the pattern HH is observed?

(c) Now suppose that p is unknown, and that we use a Beta(a, b) prior to reflect our uncertainty about p (where a and b are known constants and are greater than 2). In terms of a and b, find the corresponding answers to (a) and (b) in this setting.

7. Consider a group of n roommate pairs at Harvard (so there are $2n$ students). Each of these $2n$ students independently decides randomly whether to take Stat 110, with probability p of "success" (where "success" is defined as taking Stat 110).

Let N be the number of students among these $2n$ who take Stat 110, and let X be the number of roommate pairs where both roommates in the pair take Stat 110. Find $E(X)$ and $E(X|N)$.

Stat 110 Homework 9 Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. Let X and Y be independent, positive r.v.s. with finite expected values.

(a) Give an example where $E(\frac{X}{X+Y}) \neq \frac{E(X)}{E(X+Y)}$, computing both sides exactly.
Hint: start by thinking about the simplest examples you can think of!

As a simple example, let X take on the values 1 and 3 with probability 1/2 each, and let Y take on the values 3 and 5 with probability 1/2 each. Then $E(X)/E(X+Y) = 2/(2+4) = 1/3$, but $E(X/(X+Y)) = 31/96$ (the average of the 4 possible values of $X/(X+Y)$, which are equally likely). An even simpler example is to let X be the constant 1 (a degenerate r.v.), and let Y be 1 or 3 with probability 1/2 each. Then $E(X)/E(X+Y) = 1/(1+2) = 1/3$, but $E(X/(X+Y)) = 3/8$.

(b) If X and Y are i.i.d., then is it necessarily true that $E(\frac{X}{X+Y}) = \frac{E(X)}{E(X+Y)}$?

Yes, since by symmetry $E(\frac{X}{X+Y}) = E(\frac{Y}{X+Y})$ and by linearity

$$E\left(\frac{X}{X+Y}\right) + E\left(\frac{Y}{X+Y}\right) = E\left(\frac{X+Y}{X+Y}\right) = 1,$$

so $E(\frac{X}{X+Y}) = 1/2$, while on the other hand

$$\frac{E(X)}{E(X+Y)} = \frac{E(X)}{E(X)+E(Y)} = \frac{E(X)}{E(X)+E(X)} = 1/2.$$

(c) Now let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$. Show *without using calculus* that $E(\frac{X^c}{(X+Y)^c}) = \frac{E(X^c)}{E((X+Y)^c)}$ for every real $c > 0$.

The equation we need to show can be paraphrased as the statement that $X^c/(X+Y)^c$ and $(X+Y)^c$ are uncorrelated. As shown in class in the bank-post office story, $X/(X+Y)$ is independent of $X+Y$. So $X^c/(X+Y)^c$ is independent of $(X+Y)^c$, which shows that they are uncorrelated.

2. The *Gumbel distribution* is the distribution of $-\log X$ with $X \sim \text{Expo}(1)$.

(a) Find the CDF of the Gumbel distribution.

Let G be Gumbel and $X \sim \text{Expo}(1)$. The CDF is

$$P(G \leq t) = P(-\log X \leq t) = P(X \geq e^{-t}) = e^{-e^{-t}}$$

for all real t.

(b) Let $X_1, X_2, \dots$ be i.i.d. Expo(1) and let $M_n = \max(X_1, \dots, X_n)$. Show that $M_n - \log n$ converges in distribution to the Gumbel distribution, i.e., as $n \to \infty$ the CDF of $M_n - \log n$ converges to the Gumbel CDF.

The CDF of $M_n - \log n$ is

$$P(M_n - \log n \leq t) = P(X_1 \leq t + \log n, \dots, X_n \leq t + \log n) = P(X_1 \leq t + \log n)^n.$$

Using the Expo CDF and the fact that $(1 + \frac{x}{n})^n \to e^x$ as $n \to \infty$, this becomes

$$(1 - e^{-(t+\log n)})^n = (1 - \frac{e^{-t}}{n})^n \to e^{-e^{-t}}.$$

3. Consider independent Bernoulli trials with probability p of success for each. Let X be the number of failures incurred before getting a total of r successes.

(a) Determine what happens to the distribution of $\frac{p}{1-p}X$ as $p \to 0$, using MGFs; what is the PDF of the limiting distribution, and its name and parameters if it is one we have studied?

Hint: start by finding the Geom(p) MGF. Then find the MGF of $\frac{p}{1-p}X$, and use the fact that if the MGFs of r.v.s Y_n converge to the MGF of a r.v. Y, then the CDFs of the Y_n converge to the CDF of Y.

Let $q = 1 - p$. For $G \sim \text{Geom}(p)$, the MGF is

$$E(e^{tG}) = p \sum_{k=0}^{\infty} e^{tk} q^k = p \sum_{k=0}^{\infty} (qe^t)^k = \frac{p}{1 - qe^t},$$

for $qe^t < 1$. So the NBin(r, p) MGF is $\frac{p^r}{(1-qe^t)^r}$ for $qe^t < 1$. Then the MGF of $\frac{p}{1-p}X$ is

$$E(e^{\frac{tp}{q}X}) = \frac{p^r}{(1 - qe^{tp/q})^r}$$

for $qe^{tp/q} < 1$. Let us first consider the limit for $r = 1$. As $p \to 0$, the numerator goes to 0 and so does the denominator (since $qe^{tp/q} \to 1e^0 = 1$). By L'Hôpital's Rule,

$$\lim_{p \to 0} \frac{p}{1 - (1-p)e^{tp/(1-p)}} = \lim_{p \to 0} \frac{1}{e^{tp/(1-p)} - (1-p)t\left(\frac{1-p+p}{(1-p)^2}\right)e^{tp/(1-p)}} = \frac{1}{1-t}.$$

So for any fixed $r > 0$, as $p \to 0$ we have

$$E(e^{\frac{tp}{q}X}) = \frac{p^r}{(1 - qe^{tp/q})^r} \to \frac{1}{(1-t)^r}.$$

This is the Gamma$(r,1)$ MGF for $t < 1$ (note also that the condition $qe^{tp/q} < 1$ is equivalent to $t < -\frac{1-p}{p}\log(1-p)$, which converges to the condition $t < 1$ since again by L'Hôpital's Rule, $\frac{-p}{\log(1-p)} \to 1$). Thus, the scaled Negative Binomial $\frac{p}{1-p}X$ converges to Gamma$(r,1)$ in distribution as $p \to 0$.

(b) Explain intuitively why the result of (a) makes sense.

The result of (a) makes sense intuitively since the Gamma is the continuous analogue of the Negative Binomial, just as the Exponential is the continuous analogue of the Geometric (as discussed in class and seen on the HW 5 problem about *The Winds of Winter*). In fact, that problem essentially shows that if $X \sim$ Expo(1) then $\lfloor \frac{1}{\log(\frac{1}{1-p})} X \rfloor \sim$ Geom(p). To convert from discrete to continuous, imagine performing many, many trials where each is performed very, very quickly and has a very, very low chance of success. To balance the rate of trials with the chance of success, we use the scaling $\frac{p}{q}$ since this makes $E(\frac{p}{q}X) = r$, matching the Gamma$(r,1)$ mean.

4. (a) If X and Y are i.i.d. continuous r.v.s with CDF $F(x)$ and PDF $f(x)$, then $M = \max(X,Y)$ has PDF $2F(x)f(x)$. Now let X and Y be discrete and i.i.d., with CDF $F(x)$ and PMF $f(x)$. Explain in words why the PMF of M is *not* $2F(x)f(x)$.

The PMF is not $2F(x)f(x)$ in the discrete case due to the problem of ties: there is a nonzero chance that $X = Y$. We can write the PMF as $P(M = a) = P(X = a, Y < a) + P(Y = a, X < a) + P(X = Y = a)$ since $M = a$ means that at least one of X, Y equals a, with neither greater than a. The first two terms together become $2f(a)P(Y < a)$, but the third term may be nonzero and also $P(Y < a)$ may not equal $F(a) = P(Y \le a)$.

(b) Let X and Y be independent Bernoulli(1/2) r.v.s, and let $M = \max(X,Y)$, $L = \min(X,Y)$. Find the joint PMF of M and L, i.e., $P(M = a, L = b)$, and the marginal PMFs of M and L.

In order statistics notation, $L = X_{(1)}, M = X_{(2)}$. Marginally, we have $X_{(1)} \sim$ Bern(1/4), $X_{(2)} \sim$ Bern(3/4). The joint PMF is

$$\begin{aligned}
P(X_{(1)} = 0, X_{(2)} = 0) &= 1/4 \\
P(X_{(1)} = 0, X_{(2)} = 1) &= 1/2 \\
P(X_{(1)} = 1, X_{(2)} = 0) &= 0 \\
P(X_{(1)} = 1, X_{(2)} = 1) &= 1/4
\end{aligned}$$

Note that these are nonnegative and sum to 1, and that $X_{(1)}$ and $X_{(2)}$ are dependent; how does this relate to the problem of the probability of both of two children being girls, given that at least one is a girl?

5. Let $X \sim \text{Bin}(n, p)$ and $B \sim \text{Beta}(j, n-j+1)$, where n is a positive integer and j is a positive integer with $j \leq n$. Show using a story about order statistics (without using calculus) that

$$P(X \geq j) = P(B \leq p).$$

This shows that the CDF of the continuous r.v. B is closely related to the CDF of the discrete r.v. X, and is another connection between the Beta and Binomial.

Let $U_1, \dots, U_n$ be i.i.d. Unif$(0, 1)$. Think of these as Bernoulli trials, where U_j is defined to be "successful" if $U_j \leq p$ (so the probability of success is p for each trial). Let X be the number of successes. Then $X \geq j$ is the same event as $U_{(j)} \leq p$, so $P(X \geq j) = P(U_{(j)} \leq p)$.

6. A coin with probability p of Heads is flipped repeatedly. For Parts (a) and (b), suppose that p is a known constant, with $0 < p < 1$.

(a) What is the expected number of flips until the pattern HT is observed?

This can be thought of as "Wait for Heads, then wait for the first Tails after the first Heads," so the expected value is $\frac{1}{p} + \frac{1}{q}$, with $q = 1 - p$.

(b) What is the expected number of flips until the pattern HH is observed?

Let X be the waiting time for HH and condition on the first toss, writing H for the event that the first toss is Heads and T for the complement of H:

$$E(X) = E(X|H)p + E(X|T)q = E(X|H)p + (1 + EX)q.$$

To find $E(X|H)$, condition on the second toss:

$$E(X|H) = E(X|HH)p + E(X|HT)q = 2p + (2 + EX)q.$$

Solving for $E(X)$, we have

$$E(X) = \frac{1}{p} + \frac{1}{p^2}.$$

As a check, note that reduces to 6 when $p = 1/2$, which agrees with the HH vs. HT example done in class.

(c) Now suppose that p is unknown, and that we use a Beta(a, b) prior to reflect our uncertainty about p (where a and b are known constants and are greater than 2). In terms of a and b, find the corresponding answers to (a) and (b) in this setting.

Let X and Y be the number of flips until HH and until HT, respectively. By (a), $E(Y|p) = \frac{1}{p} + \frac{1}{1-p}$. So $E(Y) = E(E(Y|p)) = E(\frac{1}{p}) + E(\frac{1}{1-p})$. Likewise, by (b), $E(X) = E(E(X|p)) = E(\frac{1}{p}) + E(\frac{1}{p^2})$. By LOTUS,

$$E\left(\frac{1}{p}\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 p^{a-2}(1-p)^{b-1}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a-1)\Gamma(b)}{\Gamma(a+b-1)} = \frac{a+b-1}{a-1},$$
$$E\left(\frac{1}{1-p}\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 p^{a-1}(1-p)^{b-2}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a)\Gamma(b-1)}{\Gamma(a+b-1)} = \frac{a+b-1}{b-1},$$
$$E\left(\frac{1}{p^2}\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 p^{a-3}(1-p)^{b-1}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a-2)\Gamma(b)}{\Gamma(a+b-2)} = \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}.$$

Thus,

$$E(Y) = \frac{a+b-1}{a-1} + \frac{a+b-1}{b-1},$$
$$E(X) = \frac{a+b-1}{a-1} + \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}.$$

7. Consider a group of n roommate pairs at Harvard (so there are $2n$ students). Each of these $2n$ students independently decides randomly whether to take Stat 110, with probability p of "success" (where "success" is defined as taking Stat 110).

Let N be the number of students among these $2n$ who take Stat 110, and let X be the number of roommate pairs where both roommates in the pair take Stat 110. Find $E(X)$ and $E(X|N)$.

Create an indicator r.v. I_j for the jth roommate pair, equal to 1 if both take Stat 110. The expected value of such an indicator r.v. is p^2, so $E(X) = np^2$ by symmetry and linearity. Similarly, $E(X|N) = nE(I_1|N)$. We have

$$E(I_1|N) = \frac{N}{2n}\frac{N-1}{2n-1}$$

since given that N of the $2n$ students take Stat 110, the probability is $\frac{N}{2n}$ that any particular student takes Stat 110 (the p no longer matters), and given that one particular student in a roommate pair takes Stat 110, the probability that the other roommate does is $\frac{N-1}{2n-1}$. Or write $E(I_1|N) = \frac{\binom{N}{2}}{\binom{2n}{2}}$, since given N, the number of students in the first roommate pair who are in Stat 110 is Hypergeometric! Thus,

$$E(X|N) = nE(I_1|N) = \frac{N(N-1)}{2}\frac{1}{2n-1}.$$

Historical note: an equivalent problem was first solved in the 1760s by Daniel Bernoulli, a nephew of Jacob Bernoulli, for whom the Bernoulli distribution is named.