

1. In method 1, we measure the two sticks (of proper length  $a$  and  $b$ ). Let's call the r.v.s of those measurements  $A$  and  $B$  respectively.  $E(A) = a$  and  $E(b) = b$  and both have variances of  $\sigma^2$ . This was simple. Method 2, not so simple. First, we put the sticks end to end and take a measurement of that total length. I will call the r.v. for this measurement  $X$ . Next, we take the two sticks and set them side by side so we can measure the difference in their lengths. I will call this measurement's r.v.  $Y$ . Since the proper length of the two sticks together is  $a+b$ ,  $E(X) = a + b$ . Similarly,  $E(Y) = a - b$ . In order to actually figure out the length of the sticks, we say the stick with proper length  $a$  has measured length  $\frac{X+Y}{2}$ . Similarly, the final measurement of the stick with proper length  $b$  can be calculated as  $\frac{X-Y}{2}$ . The expected values for these do, in fact, turn out to be  $a$  and  $b$  since

$$E\left(\frac{X+Y}{2}\right) = \frac{E(X) + E(Y)}{2} = \frac{a+b + a-b}{2} = a$$

and

$$E\left(\frac{X-Y}{2}\right) = \frac{E(X) - E(Y)}{2} = \frac{a+b - (a-b)}{2} = b$$

Obviously, the means are all the same, so the methods do not differ at all so far. The difference comes in the variance.

The variance of our measurement for the stick of proper length  $a$  is  $Var\left(\frac{X+Y}{2}\right) = \frac{1}{4}(Var(X) + Var(Y) + 2Cov(X, Y))$ . Since all measurements are independent (as decided in Joe's email), we know that  $Cov(X, Y) = 0$ . This means that  $Var\left(\frac{X+Y}{2}\right) = \frac{1}{4}(Var(X) + Var(Y))$  which is simply

$$Var\left(\frac{X+Y}{2}\right) = \frac{\sigma^2}{2}$$

which is also true for  $Var\left(\frac{X-Y}{2}\right)$  as well. This way we see that the means of our measurements are the same for both methods, however, method 2 gives half the variance on both measurements and is thus superior to method 1. I expect this is because the error of each variable measurement for  $X$  and  $Y$  can cancel out often to bring the total estimate of the proper lengths closer to the true, proper lengths.

2. To see that these two expressions are equal, we can just think of each as expressions describing events for a Poisson Process. On the left hand side, we are calculating the probability that at least  $j$  poisson events occurred in time  $t$ . This is because we have the PMF for  $k$  events happening if they

are occurring at a rate of  $\lambda t$ , and we sum from  $k = j$  to infinity. The right hand side is saying the same thing, but from the perspective of the Gamma distribution. We know that the PDF of the  $\text{Gamma}(j, \lambda)$  distribution gives us the probability density that  $j$  events happened in some time  $x$ . If we integrate these probabilities over 0 to  $t$ , we find the probability that  $j$  events happened within that time  $t$ . The only question left is why we calculated for only  $j$  events on the right hand side. The answer to that is that the integral of the Gamma distribution doesn't tell us anything about the events that occurred after the moment the  $j$  events may have occurred in that time  $t$ , which is fine since we don't care. The integral shown only describes the probability that  $j$  events occurred in time  $t$ , more may have occurred as well (which is why we sum up for at least  $j$  events on the left side).

3. a) This can be done easily with LOTUS.

$$p \sim \text{Beta}(a, b)$$

We integrate the value of  $g(x)$  ( $p^2(1-p)^2$ ) over all possible values of  $p$  (0,1).

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

We know that in general, just from the integral of the Beta PDF:

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 p^{a-1}(1-p)^{b-1} dp$$

Applying this to what we see above, we can pull out the constant  $\Gamma$  factors and replace the integral with a modified form of the general equality of the integral of the Beta PDF. This leaves us with

$$E(p^2(1-p)^2) = \frac{\Gamma(a+2)\Gamma(b+2)}{\Gamma(a+b+4)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

b) No it does not. Since the way we update a Beta distribution is just by adding our observations of success or failure into the  $a$  and  $b$  parameters, and this addition process is linear, we know that adding the data one at a time in a certain order or just adding it all at once makes no difference. The Beta still ends up the same in the end. In other words, we can just say this update is coherent, and thus unaffected by the order here. The only way the order would make a difference is if we were trying to make predictions

along the way. Then our results would be affected by the order, but that's not the question here.

c) We had  $p \sim Unif(0,1)$  which is the same as  $p \sim Beta(1,1)$ . Now after the 6 wins and 4 losses, we we update this Beta to be  $Beta(7,5)$ . This is our posterior distribution for p.

d) In this context, saying "conditional on p", says "holding p constant between trials". Since these trials are conditionally independent, the indicator of winning the first game is fully uncorrelated with winning the second. However, if we condition on historical data, which can be updated with each game, we know that the trials are indeed correlated since winning the first game allows us to update the parameter a to a+1 and thus have a higher probability of winning the second game thus creating a positive correlation.

e) With the historical data, we have our probability of winning each game as modeled by  $p \sim Beta(7,5)$ . Also, we know that asking what is the probability that the match is undecided by the fifth game is the same as asking for the probability that the first four games had 2 wins, 2 losses. Given the probability of winning to be p, we know that this is just  $P(X = 2)$  where  $X \sim Binom(4,p)$ . This would expand to  $\binom{4}{2}p^2(1-p)^2$ . This is the answer!....if we were to know p. But we don't know exactly p. We do have the distribution for it though! So we jsut need to find the expected value of this expression using our distribution for p. Wait a second, we already did that in part a! Combining this expression for the binomial distribution,  $p \sim beta(7,5)$  with our historical data, and taking the results from part a, we see that  $E(P(X = 2)) = \binom{4}{2} \frac{\Gamma(9)\Gamma(7)}{\Gamma(16)} \frac{\Gamma(12)}{\Gamma(7)\Gamma(5)}$  which can be simplified leaving

$$E(P(X = 2)) = \frac{4}{13}$$

4. a) In order to describe this distribution, it is easiest to find the CDF,  $P(M_n \leq c, M_{n+1} \leq k)$ . In order for these two arguments to be true, all n variables must be less than c and all n+1 variables must be less than k. If  $c \leq k$ , then this is simply

$$P(M_n \leq c, M_{n+1} \leq k) = F(c)^n F(k)$$

However, if  $c \geq k$ , then we need all n+1 variables to be less than c, giving:

$$P(M_n \leq c, M_{n+1} \leq k) = F(c)^{n+1}$$

b) Describing this distribution requires us to rederive the PDF for the order statistic. In order for us to calculate the  $P(X_i = c, X_j = k)$ , we need to know the probability that one of the variables is within  $dx$  of  $c$ , another is within  $dx$  of  $k$ , and the the distribution of the other order statistics work out such that there are  $i - 1$  with values less than  $c$ ,  $n - j$  with values above  $k$ , and the rest are valued between  $c$  and  $k$ . There are  $n$  variables that could be very close to  $c$  so we get a term of  $n * f(c)dx$ , then, similarly, there are  $n - 1$  variables that would be close to  $k$  so we get a term of  $(n - 1) * f(k)dx$ . As for the rest of the variables, we can just look at their placement similarly to how we deal with binomial distributions since we can deal with a variable's values being greater than or less than a certain number,  $w$ , as just a success with  $P = F(w)$ . We need  $i - 1$  to be less than  $c$  so we get  $\binom{n-2}{i-1} F(c)^{i-1}$ . Next, we need  $j - i - 1$  to be between  $c$  and  $k$  so we get  $\binom{n-i}{j-i-1} (F(k) - F(c))^{j-i-1}$ . For the remaining  $n - j$  variables being above  $k$ , we get  $\binom{n-j}{n-j} (1 - F(k))^{n-j}$ . Combining all of these and dropping the  $dx$  terms, we get

$$P(X_i = c, X_j = k) = n(n - 1)f(c)f(k) \binom{n - 2}{i - 1} F(c)^{i-1} \binom{n - i - 1}{j - i - 1} (F(k) - F(c))^{j-i-1} (1 - F(k))^{n-j}$$

(I know it hits the end of the page. It's all there though.)

5. a) Since all orderings of the  $Y$  variables are equally likely, we can just take the naive definition of probability and say that the probability that  $Y_{new}$  is between  $Y_j$  and  $Y_k$  is just the number of positions between order statistics between  $i$  and  $j$  divided by the total number of orders  $Y_{new}$  could be in.

$$P(Y_{new} \in [Y_j, Y_k]) = \frac{k - j}{n + 1}$$

b) In order to solve for an interval in terms of the order statistics, we can just plug the 95% into the equation from part a and figure out suitable  $i$  and  $j$  values.

$$\frac{k - j}{99 + 1} = 0.95$$

$$k - j = 95$$

Thus, any  $Y$  values separated by more than 95 will have an interval between them into which we are at least 95% confidence that  $Y_{new}$  will fall. For instance,  $Y_3$  and  $Y_{98}$  would work.

6. If we say that the distribution of the  $X$  variables is described by CDF  $F$ , then we know that  $F(w) = \frac{1}{2}$  iff  $w$  is the median. Then, for the samples' median, we know we can just look at the middle order statistic. If we have  $n$  order statistics, and  $n$  is odd, we know that the median order statistic is just  $X_{\frac{n+1}{2}}$ , but this fraction looks bad, so let's talk about it in terms of  $n = 2k + 1$  so the median order statistic for our samples is now  $X_{k+1}$ . We know the distribution of that order statistic is, in its CDF,  $P(X_{j+1} \leq x) = \sum_{k=j+1}^n \binom{n}{k} F(x)^k (1 - F(x))^{n-k}$ . We can use this same CDF  $F$  since these  $X$  values are pulled from the same source distribution. We want to find the median of this distribution, so we could do  $F^{-1}(\frac{1}{2})$ , but I don't want to solve for the inverse of that CDF. So, I'll just guess. Maybe, the median of these samples is the same as the median of the overall distribution,  $w$ . Let's try it. If we plug in  $w$  and get  $\frac{1}{2}$  for the CDF here, we know the two distributions have the same median.

$$P(X_{j+1} \leq w) = \sum_{k=j+1}^n \binom{n}{k} F(w)^k (1 - F(w))^{n-k}$$

We know that  $F(w) = \frac{1}{2}$  since, again, these  $X$ 's are all from the same source distribution.

$$P(X_{j+1} \leq w) = \sum_{k=j+1}^n \binom{n}{k} \frac{1}{2}^n$$

This is just the binomial distribution,  $\text{Binom}(n, \frac{1}{2})$ , which we can use to our advantage. With the binomial distribution where  $p = \frac{1}{2}$ , we know that the probability that the relevant binomially distributed variable  $Y$  is greater than some value is the same as the probability that  $n - Y$  is greater than that value. This gives  $P(Y \geq X_{j+1}) = P(n - Y \geq X_{j+1})$ . We also know these have to add up to 1 since they make up all possibilities in this probability space. Thus, we get that both are equal to  $\frac{1}{2}$  which shows that the medians of both the distribution of  $X$  and the distribution of the sample medians are equal to some identical value  $w$ .