

## Lecture 8–9

3. Let  $X$  be a single sample from a Binomial distribution  $B(n, p)$ . In each of the following four cases, decide whether there exists an unbiased estimator and justify your answer.

- (a) Assume  $n$  is known but  $p$  is unknown and we would like to estimate  $p$ .

If  $n \neq 0$  then  $\frac{X}{n}$  is an unbiased estimator for  $p$ .

$$\begin{aligned}\mathbb{E}(X) &= np \implies \\ \frac{\mathbb{E}(X)}{n} &= p \implies \\ \mathbb{E}\left(\frac{X}{n}\right) &= p \quad \checkmark\end{aligned}$$

- (b) Assume  $p$  is known but  $n$  is unknown and we would like to estimate  $n$ .

If  $p \neq 0$  then  $\frac{X}{p}$  is an unbiased estimator for  $n$ .

$$\begin{aligned}\mathbb{E}(X) &= np \implies \\ \frac{\mathbb{E}(X)}{p} &= n \implies \\ \mathbb{E}\left(\frac{X}{p}\right) &= n \quad \checkmark\end{aligned}$$

- (c) Assume  $n$  and  $p \in (0, 1)$  are both unknown and we would like to estimate  $n + p$ .

There is no unbiased estimator for  $n + p$ . The only information we have is an unbiased estimator for  $n \cdot p$  – this cannot be manipulated into an estimator for  $n + p$ .

*We'll formalise a proof in the SV.*

- (d) Assume  $n$  and  $p$  are both unknown and we would like to estimate  $n \cdot p$ .

$X$  is an unbiased estimator for  $n \cdot p$ .

$$\begin{aligned}X &\sim B(n, p) \implies \\ \mathbb{E}(X) &= np \quad \checkmark\end{aligned}$$

5. Let  $X_1, X_2 \dots$  be a sequence of independent and identically distributed samples from the discrete uniform distribution over  $\{1, 2, \dots, N\}$ . Let  $Z := \min\{i \geq 1 : X_i = X_{i+1}\}$ . Compute  $\mathbb{E}[Z]$  and  $\mathbb{E}[(Z - N)^2]$ . How can you obtain an unbiased estimator for  $N$ .

In a given sequence of random samples over a discrete uniform distribution  $\{1, 2, \dots, N\}$ , the probability that the next sample is equal to the current sample is  $\frac{1}{N}$  and is constant. Hence the first  $i$  such that  $X_i = X_{i+1}$  forms a geometric distribution with parameter  $\frac{1}{N}$ . So  $Z \sim \text{Geo}\left(\frac{1}{N}\right)$ .

$$\begin{aligned}\mathbb{E}(Z) &= \mathbb{E}\left(\text{Geo}\left(\frac{1}{N}\right)\right) \\ &= N \quad \checkmark\end{aligned}$$



$$\begin{aligned}\mathbb{E}((Z - N)^2) &= \mathbb{E}(Z^2) - 2N\mathbb{E}(Z) + N^2 \\ &= \mathbb{E}\left(\text{Geo}\left(\frac{1}{N}\right)^2\right) - 2N^2 + N^2 \\ &= \frac{2 - \frac{1}{N}}{\frac{1}{N^2}} - N^2 \\ &= N^2 - N \quad \checkmark\end{aligned}$$

$Z$  is an unbiased estimator for  $N$ :

$$\mathbb{E}(Z) = N \quad \checkmark$$

*Awesome!*

8. Let  $X_1, X_2, \dots, X_n$  be i.i.d. samples from a normal distribution  $N(\mu, \sigma^2)$ , where  $\mu$  is unknown but  $\sigma$  is known.

(a) Prove that  $Z_1 = X_1$ ,  $Z_2 = \overline{X_n}$  and  $Z_3 = \frac{(Z_1 + Z_2)}{2}$  are all unbiased estimators.

$$\begin{aligned}Z_1 &= X_1 \implies \\ \mathbb{E}(Z_1) &= \mathbb{E}(X_1) \implies \\ \mathbb{E}(Z_1) &= \mathbb{E}(X) \quad \checkmark\end{aligned}$$

So  $Z_1 = X_1$  is an unbiased estimator of  $X$ .  $\checkmark$

$$\begin{aligned}Z_2 &= \overline{X_n} \\ \mathbb{E}(Z_2) &= \mathbb{E}(\overline{X_n}) \\ &= \frac{\mathbb{E}(\sum_{i=1}^n X_i)}{n} \\ &= \frac{\sum_{i=1}^n \mathbb{E}(X_i)}{n} \\ &= \frac{n\mathbb{E}(X)}{n} \\ &= \mathbb{E}(X)\end{aligned}$$

So  $Z_2$  is an unbiased estimator for  $X$ .  $\checkmark$

$$\begin{aligned}Z_3 &= \frac{Z_1 + Z_2}{2} \\ \mathbb{E}(Z_3) &= \mathbb{E}\left(\frac{Z_1 + Z_2}{2}\right) \\ &= \frac{\mathbb{E}(Z_1) + \mathbb{E}(Z_2)}{2} \\ &= \frac{\mathbb{E}(X) + \mathbb{E}(X)}{2} \\ &= \mathbb{E}(X) \quad \checkmark\end{aligned}$$

So  $Z_3$  is an unbiased estimator for  $X$ .  $\checkmark$

(b) Which of the three estimators would you choose?

I would choose  $Z_2$  since it has the lowest mean squared error (since the estimator is unbiased the MSE is equal to the variance of  $Z_2$  – which is  $\frac{\sigma^2}{n}$ ).  $Z_2$  is therefore likely to have the least error.

*Proof that  $\text{Var}(\hat{Z}_2) < \text{Var}(\hat{Z}_3)$ ?*



Although if work to compute was a significant concern or if we didn't need a good estimate or knew that  $\text{Var}(X)$  was very low, then  $Z_1$  would also be suitable.

Good insight  
- yes.

9. Consider the following modification of the problem of estimating the population size. Instead of sampling without replacement, we sample with replacement. What is the expected number of items we need to sample until we have seen  $k$  different ID's (so that we can apply the estimators from the lectures)?

Let  $K$  be the number of samples we have to choose to see  $k$  different people if we sample from a population  $p$  without replacement.

Consider this to be a sum of  $k$  geometric distributions where the  $i^{\text{th}}$  geometric distribution has parameter  $\frac{p-i+1}{p}$ . If  $p \geq k$  then the sum of the expectations of these geometric distributions is:

$$\begin{aligned}\mathbb{E}(K) &= \sum_{i=1}^k \frac{p}{p-i+1} \\ &= p(H_p - H_{p-k})\end{aligned}$$

Where  $H_i$  is the  $i^{\text{th}}$  harmonic.

This leads to the expression:

$$\mathbb{E}(K) = \begin{cases} 0 & \text{if } k = 0 \vee k = 1 \\ p(H_p - H_{p-k}) & \text{if } 1 < k \leq p \\ \infty & \text{otherwise} \end{cases}$$



Very good.

11. Let  $X$  be a random variable such that  $\mu = \mathbb{E}[X] = \frac{1}{2}$  and  $\mathbf{V}[X] = 1$ . What can you deduce about  $\mathbb{E}[\ln(2X)]$ ?

We can deduce two things about  $\mathbb{E}[\ln(2X)]$ :

Firstly:

$$\begin{aligned}\mathbb{E}(\ln(2X)) &= \mathbb{E}(\ln(1 + (2X - 1))) \\ &= \mathbb{E}(\ln(1 + 2(X - \mathbb{E}(X)))) \\ &\approx \mathbb{E}\left(2(X - \mathbb{E}(X)) - \frac{4(X - \mathbb{E}(X))^2}{2}\right) \\ &\approx 2\mathbb{E}(X - \mathbb{E}(X)) - 2\mathbb{E}((X - \mathbb{E}(X))^2) \\ &\approx -2\mathbf{V}(X) \\ &\approx -1\end{aligned}$$

I think this works  
if  $X$  is close to  
 $\mathbb{E}(X)$  right?

Secondly:

Consider  $Z = \ln 2X$ . So  $\mathbb{E}(X) = \frac{1}{2}\mathbb{E}(e^Z)$ . Since  $e^Z$  is a strictly convex function, we can apply Jensen's inequality:

strict version of

$$\frac{1}{2}\mathbb{E}(e^Z) < \frac{1}{2}\mathbb{E}(e^Z) \implies$$

$$\frac{1}{2}e^{\mathbb{E}(\ln 2X)} < \mathbb{E}(X)$$

$$\mathbb{E}(\ln 2X) < \ln 2\mathbb{E}(X)$$

$$\mathbb{E}(\ln 2X) < \ln 1$$

$$\mathbb{E}(\ln 2X) < 0$$



Good.



## Lecture 12–13

1. Let  $X$  count the number of collisions among  $k$  independent samples from a discrete uniform distribution over  $\{1, 2, \dots, n\}$ .

(a) What is  $\mathbb{E}[X]$ ?

I will define a collision to be when a sample is on a sample where at least one other sample is. So the set  $\{1, 1, 2, 3, 4\}$  has only 1 collision and the set  $\{1, 1, 1\}$  has two collisions.

The expected number of distinct values in  $i$  samples ( $\mu_i$ ) can be formed by the recurrence relation:

$$\begin{aligned}\mu_{i+1} &= \mu_i + P(\text{new value}) \\ &= \mu_i + \frac{n - \mu_i}{n} \\ &= \left(\frac{n-1}{n}\right) \mu_i + 1\end{aligned}$$

With the boundary conditions  $\mu_0 = 0$ , this has the solution:

$$\mu_k = n \left(1 - \left(\frac{n-1}{n}\right)^k\right)$$

So the expected number of collisions  $\mathbb{E}(C)$  after  $k$  samples is the number of samples minus the number of distinct samples. This is given by:

$$\mathbb{E}(C) = k - n \left(1 - \left(\frac{n-1}{n}\right)^k\right)$$

*Can simplify with binomial expansion to get  $\mathbb{E}(C) \approx \binom{k}{2} \frac{1}{n}$ .*

(b) Prove that  $\mathbb{P}[X > 0] \approx 1 - \exp\left(-\binom{k}{2} \cdot \frac{1}{n}\right)$ .

The exact probability of having 0 collisions in  $k$  samples from a uniform distribution of size  $n$  is:

$$\frac{n \mathbf{P} k}{n^k}$$



So the exact probability of having more than 0 collisions in  $k$  samples is given by:

$$\begin{aligned}
 & 1 - \frac{n!}{n^k} \\
 &= 1 - \exp(\ln n! - \ln(n-k)! - \ln n^k) \\
 &\approx 1 - \exp(n \ln n - n - (n-k) \ln(n-k) + n - k - k \ln n) \text{ since } \ln n! \approx n \ln n - n \\
 &= 1 - \exp((n-k) \ln n - (n-k) \ln(n-k) - k) \\
 &= 1 - \exp\left((n-k) \ln \frac{n}{n-k} - k\right) \\
 &= 1 - \exp\left((n-k) \ln \left(1 + \frac{k}{n-k}\right) - k\right) \\
 &\approx 1 - \exp\left((n-k) \left(\frac{k}{n-k} - \frac{k^2}{2(n-k)^2}\right) - k\right) \text{ using the Taylor expansion } \ln(1+x) \\
 &= 1 - \exp\left(k - \frac{k^2}{2(n-k)} - k\right) \\
 &= 1 - \exp\left(-\frac{k^2}{2(n-k)}\right) \\
 &\approx 1 - \exp\left(-\frac{k(k-1)}{2n}\right) \text{ for } 1 \ll k \ll n \\
 &= 1 - \exp\left(-\binom{k}{2} \cdot \frac{1}{n}\right)
 \end{aligned}$$

✓ Good.

- (c) Describe how this could be used to obtain an estimator for the population size?  
Your estimator does not need to be unbiased.

We can perform the experiment many times with  $k$  to estimate the probability  $p$  that the number of collisions is greater than zero.

Then rearrange the expression above to give:

$$n \approx \binom{k}{2} \frac{1}{-\ln(1-p)}$$

Substitute our sample  $p$  into this expression to give an expression for  $n$ .

3. Prove that the set of random variables  $\sigma_{i,j}$  are not pairwise independent.

Yup! Or use sample mean from part (a) to get  $n \approx \frac{k^2}{\mu}$

$$\begin{aligned}
 \mathbb{E}(\sigma_{i,j}) &= \mathbb{E}\left(\sum_0^n p_n^2\right) \\
 &= \mathbb{E}\left(\frac{1}{n}\right)
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{E}(\sigma_{i,k} | \sigma_{i,j}) &= \mathbb{E}(P(X = X_i)) \\
 &= \frac{2}{n}
 \end{aligned}$$

✓

So the  $\sigma_{i,j}$  are not independent. Informally, knowing that  $\sigma_{i,j} = 1$  increases our estimate of the probability of  $X = X_i$  and so  $\mathbb{E}(\sigma_{i,k}) \geq \frac{1}{n}$ . ✓ Good.

4. Prove formally that any testing algorithm must have a two-sided error.

A testing algorithm which always says any distribution it tests is uniformly distributed would only have a one-sided error. So the statement is incorrect unless we add some



Oh yes, good point.  
An algorithm that  
uses the data must have  
two sided error.

further criteria into our definition of a test: a test must have a nonzero probability of predicting either result and must terminate for all inputs.

Consider three discrete random variables  $X, Y \sim \text{Po}(\lambda)$  and  $Z = X - Y$ . So  $Z$  can take any positive or negative value with nonzero probability. By our definition of a test, there exists some finite sequence of results for which the test will predict that a variable is a uniform variable. Since the probability of  $Z$  taking any discrete value is nonzero and  $Z$  is not affected by previous values, any finite-length discrete sequence of values has a nonzero probability of occurring. So  $Z$  has a nonzero probability of being incorrectly identified as a uniform variable in this test. An analogous argument holds for a random uniform variable  $X' \sim U(-\infty, \infty)$ .

Since our test was arbitrary, every test which always terminates and has a nonzero probability of predicting either result must have two-sided error.

✓ Good.

5. What is the expected number of local maxima in the secretary problem for  $n$  candidates? Based on this result, suggest an algorithm that outperforms the primitive approach.

The expected number of local maxima  $\mathbb{E}(M)$  is the sum of the probabilities that the current secretary is better than all previous secretaries. This is given by:

$$\begin{aligned}\mathbb{E}(M) &= \sum_{i=1}^n \frac{1}{i} \\ &= H_n \\ &\approx \ln n\end{aligned}$$

✓

Yep! We'll analyse the success  
probability of a related  
strategy in the SV.

Wait until you've seen  $\ln n - 1$  local maxima and then select the next person who beats all previous candidates – else select the last person.

7. Assume  $n = 4$  in the secretary problem, and for any  $1 \leq k \leq 4$  consider the strategy that accepts the first candidate that is better than the previous  $k - 1$  candidates. For each possible value of  $k$ , compute the probability of hiring the best candidate.

For  $k = 1$ , the probability that we choose the best candidate is the probability that the first candidate is the best candidate. This is  $\frac{1}{4}$ . ✓

For  $k = 2$ , the probability that we choose the best candidate is the probability that the best candidate is the second candidate or the best candidate is the third and the second best is either first or last or the best candidate is last and the second best was first. This is given by  $\frac{1}{4} + \frac{1}{4} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{4} = \frac{7}{16}$ . ✓

For  $k = 3$ , the probability that we choose the best candidate is the probability that the best candidate is third or the best candidate is fourth and the second best is not third. This is given by  $\frac{1}{4} + \frac{1}{4} \times \frac{2}{3} = \frac{7}{16}$ . ✗

For  $k = 4$ , the probability that we choose the best candidate is the probability that the best candidate is the fourth candidate. Which is  $\frac{1}{4}$ . ✓

8. Consider the secretary problem and let  $I_1, I_2, \dots, I_n$  be the  $n$  random variables where  $I_j = 1$  if and only if the  $j^{\text{th}}$  candidate is the best among the first  $j$  candidates. Prove that these  $n$  random variables are independent.

Prove by induction the lemma that:

Good.

$$\int_0^{y_{k+1}} \dots \int_0^{y_3} \int_0^{y_2} y_1^{x_1-1} y_2^{x_2-x_1-1} \dots y_k^{x_k-x_{k-1}-1} dy_1 dy_2 \dots dy_k = \frac{1}{\prod_{i=1}^k x_i} [y_k^{x_k}]_0^{y_{k+1}}$$



For  $k = 1$ :

$$\begin{aligned} & \int_0^{y_2} y_1^{x_1-1} dx \\ &= \frac{1}{x_1} [y_1^{x_1}]_0^{y_1} \\ &= \prod_{i=1}^1 \frac{1}{x_i} [y_1^{x_1}]_0^{y_1} \end{aligned}$$

So the lemma holds for  $k = 1$ .

Assume now that it holds for  $k = \ell$ .

So:

$$\begin{aligned} & \int_0^{y_{\ell+2}} \int_0^{y_{\ell+1}} \cdots \int_0^{y_3} \int_0^{y_2} y_1^{x_1-1} y_2^{x_2-x_1-1} \cdots y_{\ell}^{x_{\ell}-x_{\ell-1}-1} y_{\ell+1}^{x_{\ell+1}-x_{\ell}-1} dy_1 dy_2 \cdots dy_{\ell} dy_{\ell+1} \\ &= \int_0^{y_{\ell+2}} x^{x_{\ell+1}-x_{\ell}-1} \prod_{i=1}^{\ell} \frac{1}{x_i} [y_{\ell}^{x_{\ell}}]_0^{y_{\ell+1}} dy_{\ell+1} \text{ by induction hypothesis} \\ &= \frac{1}{\prod_{i=1}^{\ell} x_i} \int_0^{y_{\ell+2}} y_{\ell+1}^{x_{\ell+1}-x_{\ell}-1} (y_{\ell+1}^{x_{\ell}} - 0) dy_{\ell+1} \\ &= \prod_{i=1}^{\ell} \frac{1}{x_i} \int_0^{y_{\ell+2}} y_{\ell+1}^{x_{\ell+1}-1} dy_{\ell+1} \\ &= \prod_{i=1}^{\ell+1} \frac{1}{x_i} [y_{\ell+1}^{x_{\ell+1}}]_0^{y_{\ell+2}} \text{ as required} \end{aligned}$$

Since the lemma holds for  $k = 1$  and if it holds for  $k = \ell$  then it also holds for  $k = \ell + 1$ , by induction it must hold for all  $k \in \mathbb{N}_{\geq 1}$ .

I will prove that  $\forall x_i \in \mathbb{N}. P(I_{x_k} \dots I_{x_1}) = P(I_{x_k}) \times \cdots \times P(I_{x_1})$ . Without loss of generality, consider  $x_k > x_{k-1} \dots x_1$ .

$P(I_{x_k} \dots I_{x_1})$  is equal to the probability that  $\forall i. \forall x_{i-1} < j < x_i. X_j < x_i$ . There are  $x_i - x_{i-1} - 1$  such  $X_j$ . This forms the infinite depth integral:

$$\begin{aligned} P(I_{x_k} \dots I_{x_1}) &= \int_0^1 \int_0^{y_k} \cdots \int_0^{y_2} y_k^{x_k-x_{k-1}-1} y_{k-1}^{x_{k-1}-x_{k-2}-1} \cdots y_1^{x_1-1} dy_1 \dots dy_{k-1} dy_k \\ &= \prod_{i=1}^k \frac{1}{x_i} [y_k^{x_k}]_0^1 \text{ by the lemma proved above} \\ &= \prod_{i=1}^k \frac{1}{x_i} \\ &= \prod_{i=1}^k P(I_{x_i}) \end{aligned}$$

Since the  $x_i$  were arbitrary, this holds for any set of  $x_i$ . So all  $n$  random variables are independent. ✓

*A awesome !*

