CS 70 HW #14

Sunday

Question #1 – Its Raining Fish

a) The number of fish arriving in segment $[0,n]$ of the road is sum of number of fish arriving in $[i, i+1]$, $i = 0, \ldots, n-1$

↳ if X and Y are independent and $X \sim (\text{Pois}(\lambda))$ and $Y \sim (\text{Pois}(\mu))$, then $X+Y \sim \text{Pois}(\lambda+\mu)$

↓

 # Fish in $[0,n]$ is $\text{Pois}(n\lambda)$

b) $[a,b]$ interval in $[0,1]$

↳ Poisson distribution: $\lambda(b-a)$ parameter

↳ independent

↓ distribution

$\text{Pois}((b-a)\lambda)$

c) The answer is the same as b

↳ $\text{Pois}((b-a)\lambda)$.

Nothing changes except for $[a,b]$ w/ $a \geq 0$

This tells me that if $b > 0$, then nothing will change. However, if i smallest integer such that $i \geq a$ and j the largest integer where $j \leq b$, then the distribution is

$$\text{Pois}((i-a)\lambda) + \text{Pois}((j-i)\lambda) + \text{Pois}((b-j)\lambda)$$

$$= \text{Pois}((b-a)\lambda)$$

d) the distance is expo (λ)

↳ cdf matching exponential cdf

let x: distance of the first fish from the north

with other words

↳ $$\frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

↳ we get $\Pr[X < t] = 1 - e^{-\lambda t}$

e) It wouldn't change from d → expo (λ)

we get $P[X \geq t]$

is equal to probability that no fish has

in segment $[x, x+t]$ since is equal to

$e^{-\lambda t}$ since # fish in segment is Pois (λt)

Question 2 - Noisy love

a) Probability I'm correct:

$$P(X=0)\,P(Y \le 0.5 \mid X=0) + P(X=1)\,P(Y > 0.5 \mid X=1)$$

$$= 0.7P(N(0, 0.49) \le 0.5) + 0.3P(N(1, 0.49) > 0.5)$$

$$= 0.7P\left(N(0,1) \le \frac{0.5}{0.7}\right) + 0.3P\left(N(0,1) > -\frac{0.5}{0.7}\right)$$

$$= P\left(N(0,1) \le \frac{5}{7}\right)$$

$$= \Phi\left(\frac{5}{7}\right)$$

$$\simeq 0.762$$

b) Conditioning ($Y \in [0.6, 0.6 + \delta]$)

$$P(X=1 \mid Y \in [0.6, 0.6+\delta])$$

$$= \frac{P(X=1)\, P(Y \in [0.6, 0.6+\delta] \mid X=1)}{P(X=0)\, P(Y \in [0.6, 0.6+\delta] \mid X=0) + P(Y=1)\, P \in [0.6, 0.6+\delta \mid X=1]}$$

$$= \frac{P(X=1)\, f_{Y|1}(0.6)}{P(X=0)\, f_{Y|0}(0.6) + P(X=1)\, f_{Y|1}(0.6)}$$

$f_{Y|0}$ = density Gaussian w/ mean 0 and variance 0.49

$f_{Y|1}$ = density Gaussian w/ mean 1 variance 0.49

$\downarrow$ compute

$$P(X=1 \mid Y=0.6)$$

$$= \frac{0.3\,(2\pi)^{-1/2} \exp(-0.4^2/(2 \cdot 0.49))}{0.7\,(2\pi)^{-1/2} \exp(-0.6^2/(2 \cdot 0.49)) + 0.3\,(2\pi)^{-1/2} \exp(-0.4^2/(2 \cdot 0.49))}$$

$$= 0.345$$

c) Target: $P(X=1 \mid Y \in [y, y+\delta]) \geq P(X=0 \mid Y \in [y, y+\delta])$

equivalent to: $P(X=1 \mid Y \in [y, y+\delta]) \geq 1/2$

$\hookrightarrow$ LHS of previous part

$$P(X=1 \mid Y \in [y, y+\delta]) = \frac{P(X=1)\, f_{Y|1}(y)}{P(X=0)\, f_{Y|0}(y) + P(X=1) f_{Y|1}}$$

$$\hookrightarrow = \frac{0.3 \exp(-(1-y)^2/0.98)}{0.7 \exp(-y^2/0.98) + 0.3 \exp(-(1-y)^2/0.98)}$$

$$= \frac{1}{1 + (0.7/0.3) \exp((1-y)^2 - y^2)/0.98)}$$

For the RHS $\geq 1/2$

$$\frac{0.7}{0.3} \exp\left(\frac{(1-y)^2 - y^2}{0.98}\right) \leq 1$$

$$\exp\left(\frac{(1-y)^2 - y^2}{0.98}\right) \leq 3/7$$

$$\hookrightarrow 1 - 2y \leq 0.98 \ln(3/7)$$

↳ condition

$$y \geq \frac{1}{2}\left(1 - 0.98 \ln(3/7)\right) \approx 0.915$$

↳ your loss minimized by saying you would if observe a measure that is ≥ 0.915

2) Testing from other novels

$$P(X=0)\,P(Y \leq 0.915 \mid X=0) + P(X=1)\,P(Y > 0.915 \mid X=1)$$

$$= 0.7\,P(N(0, 0.49) \leq 0.915) + 0.3\,P(N(1, 0.49) > 0.915)$$

$$= 0.7\,P\left(N(0,1) \leq \frac{0.915}{0.7}\right) + 0.3\,P\left(N(0,1) > \frac{-0.085}{0.7}\right)$$

$$= 0.7\,\Phi\left(\frac{0.915}{0.7}\right) + 0.3\,\Phi\left(\frac{0.085}{0.7}\right)$$

$$\approx 0.798$$

↳ This strategy works better

Question 3: Chernoff's Inequality v CLT

a) $E[X_i] = -1/12 + 9/12 + 4/12 = 1$

$\hookrightarrow var(X_i) = \frac{1}{12}(2^2) + \frac{9}{12}(0^2) + \frac{2}{12}(1^2) = 1/2$

→ Using linearity of expectation + variance

(X_i independent)

$\hookrightarrow E\left[\sum_{i=1}^{n} X_i\right] = n$ $\quad\hookrightarrow var\left(\sum_{i=1}^{n} X_i\right) = n/2$

$\hookrightarrow E\left[\sum_{i=1}^{n} X_i - n\right] = n - n = 0$ $\quad\hookrightarrow var\left[\sum_{i=1}^{n} X_i - n\right] = \frac{n}{2}$

↳ Using scaling properties of expectation + variance

$E[Z_n] = \frac{0}{\sqrt{n/2}} = 0$

$var(Z_n) = \frac{n}{2} \div \frac{n}{2} = 1$

b) $Pr[|Z_n| \geq 2] \leq \frac{Var(Z_n)}{2^2} = \frac{1}{4}$

c) It would be $1/4$ for both since

$Pr[Z_n \geq 2] \leq Pr[|Z_n| \geq 2]$ and $Pr[Z_n \leq -2] \leq Pr[|Z_n| \geq 2]$

d) By CLT, $Z_n \to N(0,1)$, the standard normal distribution

e) Since $Z_n \to N(0,1)$

↳ approximately $Pr[|Z_n| \geq 2] \approx 1 - 0.9545$

$= 0.0455$

↳ By symmetry of normal distribution

$Pr[Z_n \geq 2] = Pr[Z_n \leq -2] \approx \frac{0.0455}{2} = 0.02275$

Question 4 – Uniform Estimator

a) I would expect this sample mean to fail since every sample mean should be close to 0 because the samples on average cancel eachother out which doesn't explain anything about theta.

b) PDF distribution

$$P[a \leq x \leq b] = \int_a^b f(x)\,dx \quad \text{for all } a < b$$

In this case

$$\frac{1}{\theta - (-\theta)} \rightarrow \frac{1}{2\theta}$$

c) applying CLT to part B

↳ By CLT, we get the normal distribution

* If we take the average of n^2 uniform random variables, we will approach a bell shape curve distribution (gaussian) w/ mean $\theta^2/2$ w/ variance $\theta^2/4$

d) unbiased estimator for θ^2

↳ this should be the value from from var... essentially times a factor of 2

↳ from part c:

$$E[2v] \to 2\left(\frac{\theta^2}{2}\right) \to \theta^2$$

e) $\sigma^2 = \text{Var}[V_i^2]$

Confidence interval of σ; $0 < \delta < 1$

i) Sample mean $[2V] \pm \delta \left(\frac{\sigma}{\sqrt{n}} \right)$

Confidence interval

ii) Suppose n is large

↳ $[2V] \pm \delta \left(\frac{\sigma}{\infty} \right)$

↳ $\theta^2 \pm \delta \left(\frac{\theta}{\phi} \right)$

↳ CNR

↳ $\theta^2 \pm \delta \left(\frac{\sigma^2}{\sigma} \right)$

↳ inner