Homework 1

86-595 Neural Data Analysis

Name: Shashank Singh
Email: sss1@andrew.cmu.edu

Due: Tuesday, September 18, 2012

Problem 1

Suppose that $X \sim \text{Poisson}(\lambda)$, and recall that $\mathsf{E}[X] = \lambda$.

$$\begin{aligned}
\mathsf{E}\left[X^2\right] &= \sum_{i=1}^{\infty} i^2 \cdot P(X = i) = \sum_{i=1}^{\infty} i \cdot \frac{\lambda^i}{(i-1)!} e^{-\lambda} \\
&= \sum_{i=0}^{\infty} (i+1) \cdot \frac{\lambda^{i+1}}{i!} e^{-\lambda} && \text{(re-index the summation)} \\
&= \lambda \left(\sum_{i=0}^{\infty} i \cdot \frac{\lambda^i}{i!} e^{-\lambda} + \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \right) \\
&= \lambda \left(\mathsf{E}[X] + e^{\lambda} e^{-\lambda} \right) && \text{(def. of } \mathsf{E}[X]\text{, Taylor series for } e^{\lambda}\text{)} \\
&= \lambda(\lambda + 1) = \lambda^2 + \lambda.
\end{aligned}$$

Therefore,

$$\mathsf{Var}[X] = \mathsf{E}\left[X^2\right] - \mathsf{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda = \mathsf{E}[X]. \quad \blacksquare$$

Problem 2

$\forall n \geq 0$, $P(Z = n)$ is the number of ways of satisfying $X_1 + X_2 = n$, times the probability that X_1 and X_2 satisfy that equation:

$$P(Z = n) = \sum_{i=0}^{n} \binom{n}{i} P(X_1 = i \cap X_2 = n - i) = \sum_{i=0}^{n} \binom{n}{i} P(X_1 = i) P(X_2 = n - i)$$

(the binomial coefficient comes from the fact that, of the n spikes, any i may have come from the first neuron, and the second equality follows from the independence of X_1 and X_2). Therefore,

$$\begin{aligned}
P(Z = n) &= \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\lambda_1^i}{i!} e^{-\lambda_1} \right) \left(\frac{\lambda_1^{n-i}}{(n-i)!} e^{-\lambda_1} \right) \\
&= \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} \sum_{i=0}^{n} \binom{n}{i} \lambda_1^i \lambda_2^{n-i} = \frac{(\lambda_1 + \lambda_2)^n}{n!} e^{-(\lambda_1 + \lambda_2)} && \text{(by the Binomial Theorem)}
\end{aligned}$$

Therefore, $Z \sim \boxed{\text{Poisson}(\lambda_1 + \lambda_2).}$

Problem 3

Let $X \sim \text{Poisson}(\lambda)$ be the number of spikes, and let Z be the number of spikes we detect. The value of $Z|X$ is distributed binomially (for each spike, we flip a coin biased to be heads with probability 0.9, and count the number of heads). Thus, by the Law of Total Probability, $\forall n \in \mathbb{N}$, if $p = 0.9$,

$$\begin{aligned}
P(Z = n) &= \sum_{i=n}^{\infty} P(X = i) \cdot P(Z = n|X = i) && \text{(since } Z \leq X\text{)} \\
&= \sum_{i=n}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \binom{i}{n} p^n (1-p)^{i-n} \\
&= \sum_{i=n}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \frac{i!}{n!(i-n)!} p^n (1-p)^{i-n} \\
&= \sum_{i=0}^{\infty} \lambda^{i+n} e^{-\lambda} \frac{p^n (1-p)^i}{n! i!} && \text{(re-index summation)} \\
&= \frac{(p\lambda)^n}{n!} e^{-\lambda} \sum_{i=0}^{\infty} \frac{((1-p)\lambda)^i}{i!} && \text{(factor out constants)} \\
&= \frac{(p\lambda)^n}{n!} e^{-\lambda} e^{(1-p)\lambda} = \frac{(p\lambda)^n}{n!} e^{p\lambda}. && \text{(Taylor series for exponential)}
\end{aligned}$$

Therefore, $Z \sim \boxed{\text{Poisson}(p\lambda).}$

Problem 4

Since $V \sim \mathcal{N}(v + b, \sigma^2)$, $X := (V - v) \sim \mathcal{N}(b, \sigma^2)$. Thus,

$$\mathsf{E}\left[X^2\right] = \mathsf{Var}\left[X\right] + \mathsf{E}\left[X\right]^2 = \boxed{\sigma^2 + b^2.}$$

(Isn't this true of any estimator?)

Problem 5

a. $\forall t \in \mathbb{R}$, $F(t)$ is the probability that the cue has occurs before time t, so that $1 - F(t)$ is the probability that the cue occurs after time t, and $f(t)$ is the probability that the cue occurs in the infinitesimal increment of time after t. Thus, at any given moment in time t, the probability of getting the cue given that it hasn't already occured is

$$P(\text{cue now}|\text{no cue yet}) = \frac{P(\text{cue now} \wedge \text{no cue yet})}{P(\text{no cue yet})} = \frac{P(\text{cue now})}{P(\text{no cue yet})} = \frac{f(t)}{1 - F(t)}. \quad \blacksquare$$

b. The hazard function of the exponential distribution $\mathrm{Exp}(\mu)$ is

$$\lambda_e(t) = \frac{\mu e^{-\mu t}}{e^{-\mu t}} = \mu.$$

For $t \in (a, b)$, the hazard function of the uniform distribution distribution $\mathrm{U}(a, b)$ is

$$\lambda_u(t) = \frac{\left(\frac{1}{b-a}\right)}{\left(\frac{t-a}{b-a}\right)} = \frac{1}{t-a}.$$

Since λ_e is constant whereas λ_u is strictly increasing (or constant 0), the subject cannot prepare for an exponentially distributed cue time, whereas, with a uniformly distributed cue time, the subject would become increasing expectant of a cue over time (in fact, theres a 3-line proof that the exponential distribution is the *only* nice distribution with a constant hazard function). Thus, the exponential distribution is better.

Problem 6

---

a. $\rho_f \approx 0.1728$. See Figure 1.

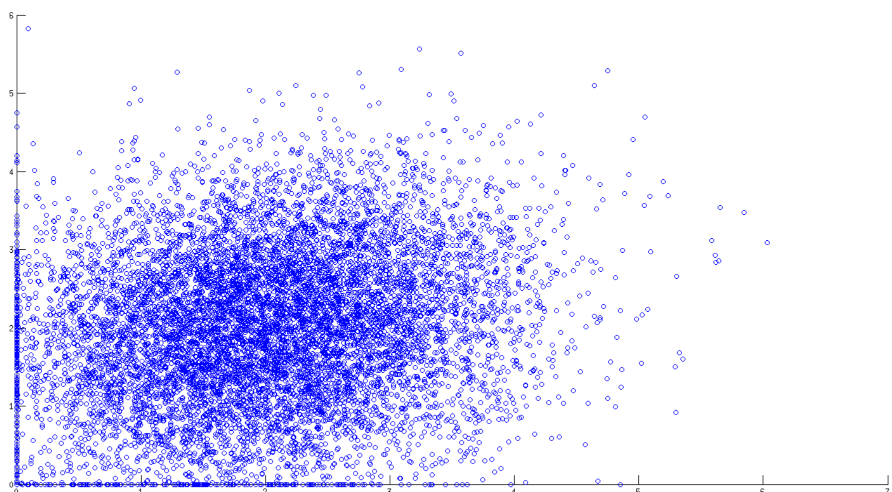

Figure 1: Scatter plot of f_2 over f_1, for $\mu = 2, \rho_v = 0.2$.

b. $\rho_f \approx 0.0203$. See Figure 2.

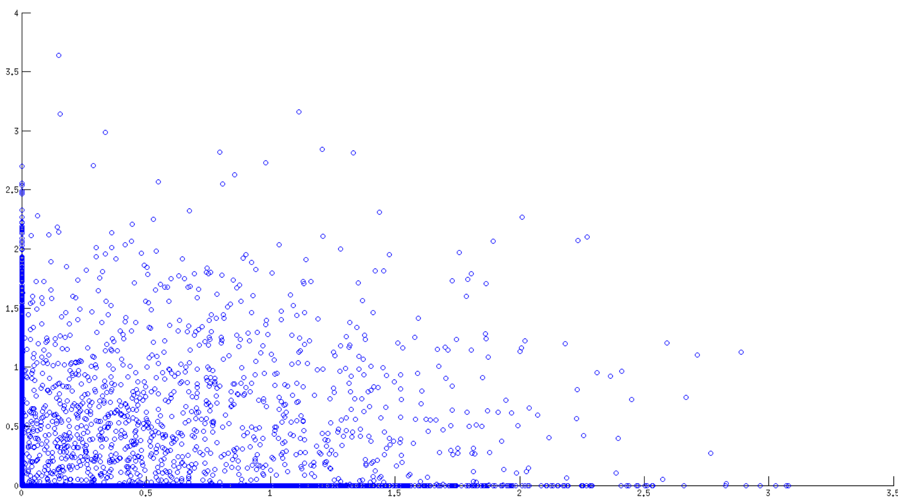

Figure 2: Scatter plot of f_2 over f_1, for $\mu = -0.5, \rho_v = 0.2$.

c. See Figure 3.

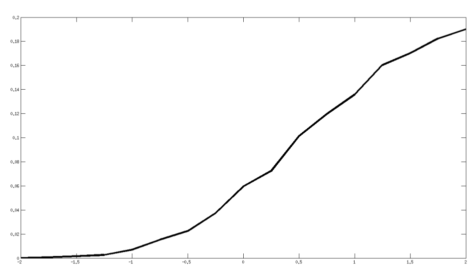

Figure 3: Plot of ρ_f as a function of μ, for $\rho_v = 0.2$.

d. See Figure 4.

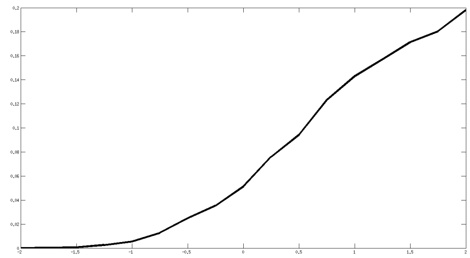

Figure 4: Plot of ρ_f as a function of μ, for $\rho_v = 0.9$.

e. See Figure 5.

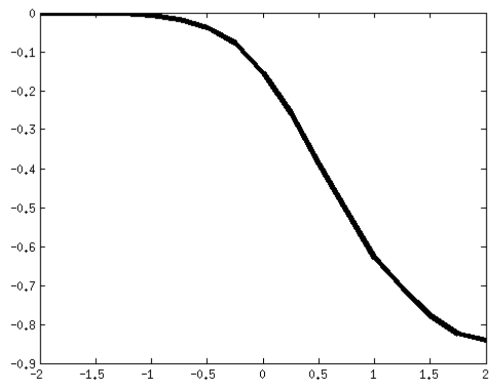

Figure 5: Plot of ρ_f as a function of μ, for $\rho_v = -0.9$.

f. Thresholding too aggressively kills correlation. As shown in Figure 2, having a threshold high relative to the mean voltages of the neurons allows only data points where both neurons displayed high voltage display correlation; if either neuron fires below the threshold rate, then the data point is pushed onto the corresponding axis. In particular, this means that, when both neurons are at low voltage (as strongly correlated neurons would often be), the data is essentially ignored.

Problem 7

I took a probabillity class last semester in which I had already done problems 1, 2, 3, and 5, so these were pretty quick. Problem 4 took me a little while to understand, but was easy once I got the setup straight, and didn't take more than 15 minutes in all. Problem 6 (the MATLAB simulation) taught me the most.