Problem 0: Readings
Read notes 4 and 5 from the course web site.

Problem 1: Poisson approximation
Assume that the production of items follows a Bernoulli scheme of trials, and the probability that an item is defective is \( p = 0.015 \). Consider boxes that contain \( n = 100 \) items.

(a) Come up with expressions for the following probabilities (you do not have to compute the actual value):

- a box contains no defective items
- a box contains 1 defective item
- a box contains 2 defective items
- a box contains \( k \) defective items

Solution:

\[
P(\text{no defects}) = \binom{100}{0} 0.015^0 (1 - 0.015)^{100}
\]

\[
P(1 \text{ defect}) = \binom{100}{1} 0.015^1 (1 - 0.015)^{99}
\]

\[
P(2 \text{ defects}) = \binom{100}{2} 0.015^2 (1 - 0.015)^{98}
\]

\[
P(k \text{ defects}) = \binom{100}{k} 0.015^k (1 - 0.015)^{100-k}
\]

(b) Using Poisson approximation, find approximate values for the above probabilities.

Solution: We compute \( \lambda = np = 100 \cdot 0.015 = 1.5 \).

\[
P(\text{no defects}) \approx \frac{1.5^0 e^{-1.5}}{0!} = 0.22313016
\]
\[ P(1 \text{ defect}) \approx \frac{1.5^1 e^{-1.5}}{1!} = 0.33469524 \]
\[ P(2 \text{ defects}) \approx \frac{1.5^2 e^{-1.5}}{2!} = 0.25102143 \]
\[ P(k \text{ defects}) \approx \frac{1.5^k e^{-1.5}}{k!} \]

(c) Assume that we are not sure about the value of \( p \), but it could be any value in the following set:
\[ \{0.010, 0.011, 0.012, 0.013, 0.014, 0.015, 0.016, 0.017, 0.018, 0.019\} \]

We open a box and find 1 defective item. What should we believe the value of \( p \) is? Explain your answer.

**Solution:** Let \( X \) be the number of defective items, which is observed. Let us also use, since we know nothing about \( p \), a uniform prior, i.e. \( P(p = v) = \frac{1}{10} \).

Let’s apply Bayes’ rule:
\[ P(p = v | X = k) = \frac{P(X = k | p = v) P(p = v)}{P(X = k)} \]

Since \( P(p = v) = \frac{1}{10} \) regardless of \( v \), and \( P(X = k) \) does not depend on \( v \), \( P(p = v | X = k) \) is maximized when \( P(X = k | p = v) \) is maximized. Now this can be approximated by \((nv)^k e^{-nv}/k!\). We seek to maximize \((nv)^k e^{-nv}\).

By taking the first derivative and setting it to zero, we find that the maximum occurs at \( v = k/n \). Therefore, the most likely value for \( p \) should be around 1/100 = 0.01. So we pick 0.01 as the most believable value for \( p \).

**Problem 2: I call this problem time warping**

Let \( f_T(t|\lambda) = \lambda e^{-\lambda t}, \ t \geq 0 \). So \( T \) is exponentially distributed when conditioned on \( \lambda \). Assume that we have a prior for \( \lambda \) that is continuous uniform between \( a \) and \( b \).

(a) Given the event \( E \) that \( T > \tau \), what can we say about \( \lambda \)? In other words, find \( f(\lambda | E) \). Recall that Bayes rule allows for mixing probabilities and densities. In particular:
\[ f(\lambda | E) = \frac{P(E | \lambda) f(\lambda)}{P(E)} \]

**Solution:**
\[ P(E | \lambda) = P(T > \tau | \lambda) = \int_{\tau}^{\infty} \lambda e^{-\lambda t} dt = e^{-\lambda \tau} \]
\[ f(\lambda) = \frac{1}{b - a} \]
\[ P(E) = \int_{a}^{b} P(E | \lambda) f(\lambda) d\lambda = \int_{a}^{b} e^{-\lambda \tau} \frac{1}{b - a} d\lambda = \frac{e^{-a \tau} - e^{-b \tau}}{\tau (b - a)} \]

Therefore,
\[ f(\lambda | E) = \frac{\tau e^{-\lambda \tau}}{e^{-a \tau} - e^{-b \tau}} \]
(b) What happens to \( f(\lambda|E) \) when we take the limit as \( a \to 0 \) and \( b \to \infty \)?

**Solution:** We get:

\[
f(\lambda|E) = \tau e^{-\tau \lambda}
\]

I call this time warping, because the role of time and rate are exchanged, i.e. \( \lambda \) is now exponentially distributed with rate \( \tau \). This happens if we “know nothing” (improper uniform prior) about \( \lambda \) except that it is positive.

**Problem 3: Central Limit**

Consider tossing a fair coin 100\( n \) times, where \( n \) is geometrically distributed with mean 6. Due to a counting error, it is reported that the fraction of heads is in \([0.5, 0.51]\). What is the most likely value for \( n \)?

**Solution:** Let \( S_{100n} \) be the number of heads obtained in 100\( n \) tosses. We are given that 50\( n \leq S_{100n} \leq 51n \). We can write:

\[
P(50 \leq S_{100n} \leq 51\sqrt{n}|n) = P(0 \leq S_{100n} - 50n \leq \sqrt{n}|n) = P(0 \leq \frac{S_{100n} - 50n}{0.5\sqrt{100n}} \leq 0.2\sqrt{n}|n)
\]

By the central limit theorem, \((S_{100n} - 50n)/(0.5\sqrt{100n})\) converges in distribution to standard Normal (with mean 0 and variance 1). So

\[
P(0 \leq \frac{S_{100n} - 50n}{0.5\sqrt{100n}} \leq 0.2\sqrt{n}|n) = \Phi(0.2\sqrt{n}) - \Phi(0.2\sqrt{n}) - 1/2
\]

Now,

\[
P(n|50n \leq S_{100n} \leq 51n) \propto P(50n \leq S_{100n} \leq 51n|n)P(n)
\]

\[
\propto [\Phi(0.2\sqrt{n}) - 1/2](5/6)^{n-1}1/6
\]

For various values of \( n \), we have:

\[
\begin{align*}
n &= 1 : (0.5793 - 0.5) = 0.0793 \\
n &= 2 : (0.6103 - 0.5)(5/6) = 0.0919 \\
n &= 3 : (0.6368 - 0.5)(5/6)^2 = 0.095 \\
n &= 4 : (0.6554 - 0.5)(5/6)^3 = 0.0899 \\
n &= 5 : (0.6736 - 0.5)(5/6)^4 = 0.0837
\end{align*}
\]

So the most likely value of \( n \) is 3.