Exercises: Distributions

1. Binomial

A sensor in a smart network has a probability $p = .7$ to correctly detect a person entering a room.

a) If 10 people enter the room, how many detections are expected?

$E[X] = \mu = np = 10 * 0.7 = 7$

b) If 12 people enter the room, what is the probability that between 2 and 7 people are detected?

$P[X \leq 7] = 0.2763$, $P[X \leq 1] = 0$, *from table or python,* $P[2 \leq X \leq 7] = 0.2763 - 0$ *, ie the probability is .28 or 28%*

c) If 3 people are detected, what's the likelihood that 5 people entered the room?

$\mathcal{L}(p|n, x) = \binom{n}{x} p^x (1-p)^{n-x}$ *for $n = 5, p = .7$ which gives $l \simeq 0.309$. Ie, about a 30% probability that an observation of 3 was drawn from a* $\mathrm{Binom}(n = 5, p = .7)$ *distribution. Note that the technical question answered is "given 3 detections and 5 people entering the room, what is the likelihood of $X \sim \mathrm{Binom}(n = 5, p = .7)$", we draw conclusions since the distribution is known! Note the obvious trap; for $p = .6$ the likelihood is $l = .3456$ and it could be even higher for some other distribution with completely different parameters– so likelihood isn't **evidence** of anything! It only indicates the degree of support for some **assumption** (specifically the conditionals of the likelihood function).*

2. Standard Normal

A large random sample from an unknown distribution is drawn. The mean is calculated to be 0 and the standard deviation is 1.

a) What kind of distributions could the sample have been drawn from? *Only a normal distribution, since $\frac{X-\mu}{\sigma}$ is just X and fully described by $X \sim N(0, 1) = Z$.*

b) What's the probablility that a normally distributed random variable will lie within one, two and three standard deviations from its mean?

Consider the smart network sensors in the previous exercise. Across 120 common rooms, a total of 24000 people move through an office complex per day.

c) How many detections are expected per day?

$24000 * 0.7 = 16800$ *as before.*

d) How many detections would be considered unusually few?

$\sigma = \sqrt{npq} \simeq 71$ *so less than 16500 detections would be quite few.*

3. Geometric

During a heavy traffic scenario, it was found that the probability of successfully logging on to a network router at any point was 70%.

a) Is it unusual if two attempts are required to log in during heavy traffic?

No. The mean is ~ 1.43 and $\sigma \simeq 0.78$, ie 68% of values are in the range $[0.65, 2.21]$, requiring one or two attempts.

b) What is the likelihood of observing (exactly) three attempts?

*$\mathcal{L}(p|x) = (1-p)^x p$, or 1.89%. It's certainly unusual, but not quite on rare ($3\sigma \sim 99.7\%$). Note that this technically only indicates agreement with a geometric distribution (or lack thereof) for these specific data points. Since the problem has a known distribution, we can infer the conclusion; unusual but not quite rare. The "backwards" nature of likelihood means that the conditional probability is actually "Given that it took three attempts, what's the likelihood this was drawn from a gemoetric distribution with $p = .7$ ", ie the preceding argument is a kind of pseudo-Bayesian reasoning. It's **not** well-formed and there are cases where real applications of Bayes is needed (not for Geom though). Tread carefully!*

c) At what point does the number of attempts become unusual?

About 16% of expected values lie between $1 - 3\sigma$. In the upper range that is considered unusual and beyond that rare, so three attempts is unusual and four or more is rare.

*Tutors note: These are **statistical** arguments relying on the Central Limit Theorem, c.f. probabilistic arguments, $P[X > 1] = 0.3$ ie 30% of logins during heavy traffic obviously require more than one attempt and $f(3) = .063$, ie for any login there's a 6.3% probability that three attempts are needed.*

4. Negative binomial

As a follow-up to the heavy traffic scenario, a system was devised that would automatically log in to a number of routers in order to retain monitoring and control in case of network congestion. 12 routers are part of the system. The logins are independent of each other. The probability of a successful login attempt remains as $p = .7$.

a) How many attempts are expected before all 12 systems are connected?

$E[X] = r/p = 12/.7 = 17.14$. In scipy neg-binom counts number of failures, not number of attempts, so add r!

b) What is the probability of exactly needing 12 attempts?

$f(12) = 0.0138$(in scipy: `.pmf(0)`) ie 1.38%

c) In a reasonable worst-case scenario, how many attempts could be expected to be required?

$\sigma = \sqrt{rq/p^2} \simeq 7.35$ so 39 attempts are 3σ from the mean. Ie, in a reasonable worst-case scenario 40 attempts can be expected before all 12 routers are connected.