School of Mathematics and Physics, UQ

STAT2003 Mathematical Probability
Semester 1 2025
Problem Set 2

Michael Kasumagic, 44302669
Tutorial Group #10
Due 1pm Monday 28 April 2025

Question 1

You are at a small pond that contains a total of $N = 20$ fish. Of these $N_1 = 14$ are catfish and $N_2 = 6$ are bass. You go fishing and you use bait and a fishing technique that attracts both types of fish in the same manner.

(a) You catch $n = 4$ fish without replacement. What is the $\mathbb{P}(\#\text{Catfish} = 2)$? Use a first principles counting argument.

(b) Answer the problem using the hyper-geometric distribution.

(c) Suppose you now return the fish to the pond after catching it. You catch $n = 4$ fish in total. What is the $\mathbb{P}(\#\text{Catfish} = 2)$ now?

(d) Compare the results of the two cases above. Why are the results different?

(e) The results do not differ significantly. Explain this by presenting a derivation of the binomial(n, p) distribution is the limit of the hypergeometric distribution, where we sample n elements from a population that has of N_1 of one type and and N_2 of another type. In particular assume that,

$$\lim_{N_1, N_2 \to \infty} \frac{N_1}{N_1 + N_2} = p.$$

With such a limit, show that the probability mass function of a bionomial is the limit of the hypergeometric probability mass function.
Argue what happens to $\mathbb{P}(X = x)$ as $N_1, N_2 \to \infty$.

(f) For the case of $n = 4$, and $p = 14/20$, plot the CDF of the binomial, together with the CDF of the associated hypergeometric distributions having $N_1 + N_2 = 10$, $N_1 + N_2 = 20$, and $N_1 + N_2 = 100$. Use Python for this and present your code. Explain what you see in the plot.

Solution:

Question 2

Assume a mixture distribution is parameterized by some scalar parameter θ, and has a probability density or probability mass function $f(x;\theta)$. It can also have other parameters not included in θ. Treat the parameter θ as a random variable with some known continuous probability density function $g(\cdot)$. Then we have the mixture distribution (probability density or probability mass function),

$$f(x) = \int_{-\infty}^{\infty} f(x;\theta)g(\theta)\,d\theta.$$

(a) Assume $\alpha > 0$ and $\lambda > 0$ are some fixed values, and that θ is distributed as $\text{Gamma}(\alpha, \beta)$ with density,

$$g(\theta) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta}\mathbf{1}_{\{\theta>0\}}$$

Further assume that $f(x;\theta)$ is Poisson where θ is the mean. Determine the resulting mixture distribution and express it's parameters in terms of α and β. Is it a negative binomial distribution?

(b) Assume that θ is distributed as $\text{beta}(\alpha, \beta)$ for some fixed $\alpha > 0$ and $\beta > 0$, with density,

$$g(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\mathbf{1}_{\{0\le\theta\le1\}}.$$

Assume that $f(x;\theta)$ is binomial with a fixed number of trials n and the success probability parameter being θ. Determine the probability mass function of the mixture distribution. Try to represent the probability mass function using only expressions involving gamma functions $\Gamma(\cdot)$. Identify the name of this probability distribution using the distributions table supplied in the lecture.

(c) Assume that the density of θ on $u \in [0,1]$ is $g(u) = 6u(1-u)$, and is 0 elsewhere. Continuing with $f(x;\theta)$ being binomial with a fixed number of trials n and the success probability parameter being θ, determine the mixture distribution. Plot the probability mass function when $n = 20$. Present your code.

Solution:

Question 3

Consider a Pareto distribution with probability density function (PDF) given by:

$$f(x) = \frac{\alpha}{x^{\alpha+1}} \mathbf{1}_{\{x \geq 1\}},$$

where $\alpha > 0$ is the shape parameter.

1. Compute the cumulative distribution function, $F(x)$.
2. Determine for which integer values of α the r-th moment $\mathbb{E}[X^r]$ is finite.
3. A relationship that holds for non-negative continuous random variables X is,

$$\mathbb{E}[X] = \int_0^\infty \big(1 - F(x)\big)\, dx.$$

   Prove this by considering $F(x)$ as an integral and manipulating the order of integration of the double integral.

4. Use the above result to compute the mean of the Pareto distribution, and verify this against a straightforward mean computation. What is the range of the values of α for which the mean is finite?
5. Determine $F^{-1}(u)$, the inverse cdf.
6. Use the inverse transform sampling to generate 10^6 random variables for the case of $\alpha = 1.5$. Use the sample mean to estimate the mean and compare to the theoretical result.
7. **(*STAT7003)** Try to repeat for $\alpha = 1$. Generate an array of 10^6 random variables, and create a table of your sample mean estimates for the first $n = 10^5, 2\times10^5, 3\times10^5, \ldots, 10^6$ from that array. Does the estimate appear to converge? Explain your results.

Solution:

Question 4

Solution:

Question 5

Solution:

Question 6

Solution:

Question 7

Solution:

Question 8

Solution: