# MANDATORY ASSIGNMENT 2
# STK1100

IVAR STANGEBY

**Problem 1.** *We let $X$ denote the yearly salary for a person chosen at random. We further assume that $X$ is Pareto distributed. The Pareto probability distribution has the density function:*

$$f_X(x) = \begin{cases} \theta\kappa^\theta x^{-\theta-1} & \text{if } x > \kappa \\ 0 & \text{otherwise.} \end{cases}$$

*The quantity $\kappa$ is the minimum salary, where as $\theta > 2$ is a parameter dependent on the differences in salary.*

**Solution 1.**
a). We want to find the cumulative probability distribution $F_X(x)$ of $X$. This is defined as

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)\,dy,$$

however since $f_X(y)$ contributes nothing to the integral when $y$ is in the range $(-\infty, \kappa)$. We can therefore evaluate the integral

$$\begin{aligned} F_X(x) &= \int_{\kappa}^{x} f_X(y)\,dy. \\ &= \theta\kappa^\theta \int_{\kappa}^{x} y^{-\theta-1}\,dy \\ &= \theta\kappa^\theta \Big| -x^{-\theta} \Big|_{\kappa}^{x} \\ &= \kappa^\theta\kappa^{-\theta} - \kappa^\theta x^{-\theta} \\ &= 1 - \kappa^\theta x^{-\theta}, \end{aligned}$$

as required.

The *median yearly salary* is the real value $m$ that satisfies

$$F_X(m) = \frac{1}{2}.$$

Solving this equation for $m$ gives

$$m = \sqrt[\theta]{2}\kappa.$$

b). We want to find the expected value of $X$. This is given by

$$\mathbb{E}[X] = \mu_X = \int_{\kappa}^{\infty} x f_X(x)\,dx = \lim_{b\to\infty} \int_{\kappa}^{b} x f_X(x)\,dx$$

Evaluating the last integral gives us

$$\lim_{b\to\infty} \theta\kappa^\theta \Big| \frac{x^{-\theta+1}}{-\theta+1} \Big|_{\kappa}^{b},$$

which in turn equals

$$\mathbb{E}[X] = \frac{\theta\kappa}{\theta-1}$$

c). Given $\kappa = 200000$ and $\theta = 2.5$ we evaluate the median yearly salary and the expected yearly salary:

$$m = \sqrt[\theta]{2}\kappa \approx 263901,$$

$$\mathbb{E}[X] = 333333,333\ldots.$$

The median yearly salaray better captures the skewness in the salary distribution across a population, where as the expected value is biased towards the low amount of people earning most of the money.

d). We want to find the variance and standard devation of $X$. These are given by

$$\mathbb{V}[X] = \mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left[X^2\right] - \mu_X^2,$$

$$\sigma = \sqrt{\mathbb{V}[X]}.$$

The variance is easily derived as

$$\mathbb{V}[X] = \lim_{b\to\infty}\int_\kappa^b x^{-\theta+1}\,dx - \mu_X^2.$$

Solving this integral yields:

$$\mathbb{V}[X] = \frac{\kappa^2\theta}{\theta-2} - \mu_X^2 = \frac{\theta\kappa^2}{(\theta-2)(\theta-1)^2}$$

$$\sigma = \frac{\sqrt{\theta}\kappa}{\sqrt{(\theta-2)}(\theta-1)}$$

e). If we let $Y = \theta\ln(X/\kappa)$ and we solve for $X$ we get

$$X = \kappa e^{y/\theta}.$$

Substituting this into the probability density function for $X$, we get the following density function:

$$f_Y(y) = \theta\kappa^\theta\left(\kappa e^{y/\theta}\right)^{-\theta-1},$$

which simplifies to

$$f_Y(y) = \frac{\theta}{\kappa}e^{-(1+\frac{1}{\theta})y}, y > 0.$$

which we recognice as the *exponential* distribution.

**Problem 2.** *We let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random stochastic variables, i.e., their probability distributions are the same. We also let $\mu = \mathbb{E}(X_i)$, and $\sigma^2 = \mathbb{V}(X_i)$. We introduce the arithmetic mean*

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i,$$

*and the standarized mean*

$$Z_n = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}}.$$

**Solution 2.**
a). We want to show that $\mathbb{E}(Z_n) = 0$ and $\mathbb{V}(Z_n) = 1$. We know that $\mathbb{E}(\overline{X_n}) = \mu$ and that $\mathbb{V}(\overline{X_n}) = \sigma^2/n$. Starting with the expected value, we substitute the expression for $Z_n$ and find:

$$\mathbb{E}(Z_n) = \frac{1}{\sigma/\sqrt{n}}\mathbb{E}(\overline{X_n}) - \frac{1}{\sigma/\sqrt{n}}\mu = 0.$$

We use the similar approach for the variance $\mathbb{V}(Z_n)$.

$$\mathbb{V}(Z_n) = \frac{1}{\sigma^2/n}\mathbb{V}(\overline{X}) = 1$$

b). Starting with the uniform distribution, we have:

$$\mu = \mathbb{E}(X_i) = \int_0^1 x\,dx = \frac{1}{2}$$

$$\sigma^2 = \int_0^1 x^2\,dx - \frac{1}{4} = \frac{1}{12}$$

$$\sigma = \frac{1}{2\sqrt{3}}$$

The exponential distribution yields (using integration by parts):

$$\mu = \int_0^\infty xe^{-x}\,dx = 1$$
$$\sigma^2 = \int_0^\infty x^2e^{-x}\,dx = 2 - \mu^2 = 1$$
$$\sigma = 1$$

The bernoulli distribution with paramter $p$ has

$$\mu = p = \frac{1}{2}$$
$$\sigma^2 = p(1-p) = \frac{1}{4}$$
$$\sigma = \frac{1}{2}$$

c). According to the *Central Limit Theorem* the standarized versions of $\overline{X_n}$ will have the standard normal distribution.

d). Examining the first histogram in Figure 1 we see that we have something that looks like a normal distribution, however $n = 3$ is not sufficient to get a good approximation. The higher the $n$ the more concentrated the distribution is around $\mu$ according to the Central Limit Theorem

e), f). The probabilities of standard normal distributed variable falling in any of these intervals are given in Figure 2
g). We see from Figure 1 for $n = 10$ and 30 that the distribution looks more and more like the standarized normal distribution. This is in accordance with the Central Limit Theorem. The relative frequencies converges to the probabilities outlined in Appendix A.3.
h). Nothing new to see here, again this is in accordance with the Central Limit Theorem. The distribution converges to the standard normal distribution for increased values of $n$. It is very skewed at $n = 3$. The relative frequences are also converging to the probabilities in Figure 2.
i). Again, according to the central limit theorem - this converges to the standard normal distribution for increasing values of $n$.
j. It is definitely the uniform distribution that gives us the standarized mean that is closest the the standard normal distribution for the values $n = 3, 10, 30$. However if we let $n \to \infty$, we should see a similiar distribution for all of the standarized means.

**Problem 3.** *Letting $X_1, \dots, X_9$ be independent random variables with identical distribution $F$. The empirical mean,*

$$\overline{x} = 56.22,$$

*and the median,*

$$m(x_1, \dots, x_9) = 46,$$

*are both two natural estimations. We are going to examine these estimators by looking at their properties.*

**Solution 3.**
a). We want to find the standard error $SE$ and the expectation scewness for the two estimators. We do this using *non-parametric bootstrapping*. The empirical standard deviation $\hat{\sigma} = 42.48$. The simulations yields a standard error of $\approx 13,2$ for the median estimator, and $\approx 13,4$ for the mean estimator.

The bias of the mean is very low at $\approx -0.0150$, probably due to the fact that the variables are independent, where as the bias for the median is slightly larger at $\approx -0.1148$.
b). The estimator for the median of the distribution, as shown in Figure 5 looks the way that it does, because the median is defined as the observation that splits the whole samplie into two halfs, the greater and the lesser. The histogram of the estimator for the mean does look log-normally distributed.
c). If we now assume that the random variables are log-normally distributed, we know that the median is given by

$$e^\mu,$$

and the mean is given by

$$e^{\mu+\sigma^2/2}.$$

where $\mu$ is the location of the distribution, and $\sigma$ is the scale. We want to perform parametric bootstrapping on the random variables $X_1, \ldots, X_9$ in order for us to see whether our prediction is reasonable.

A reasonable estimator for the median is going to be

$$e^{\hat{\mu}}$$

and for the mean

$$e^{\hat{\mu}+\hat{\sigma}^2/2}$$

with $\hat{\mu} = mean(x)$ and $\hat{\sigma}^2 = SE(mediansim)^2$

The result of the parametric bootstrap is shown in Figure 5.

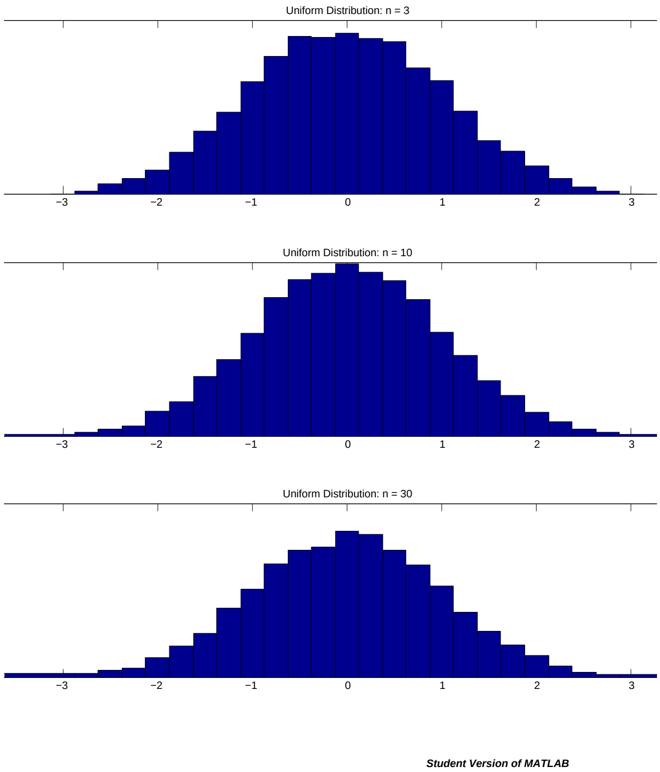


FIGURE 1. Standarized mean - Uniform Distribution

| Interval | Probability | Relative Frequency |
|---|---|---|
| $(-\infty, -2.5]$ | 0.0062 | 0.0027 |
| $(-2.5, -2.0]$ | 0.0166 | 0.0177 |
| $(-2.0, -1.5]$ | 0.0440 | 0.484 |
| $(-1.5, -1.0]$ | 0.0919 | 0.0930 |
| $(-1.0, -0.5]$ | 0.1498 | 0.1570 |
| $(-0.5, 0]$ | 0.1915 | 0.1897 |
| $[0, 0.5)$ | 0.1915 | 0.1835 |
| $[0.5, 1.0)$ | 0.1498 | 0.1454 |
| $[1.0, 1.5)$ | 0.0919 | 0.0920 |
| $[1.5, 2.0)$ | 0.0440 | 0.0517 |
| $[2.0, 2.5)$ | 0.0166 | 0.0167 |
| $[2.5, \infty)$ | 0.0062 | 0.0022 |

FIGURE 2. Probabilities and relative frequencies

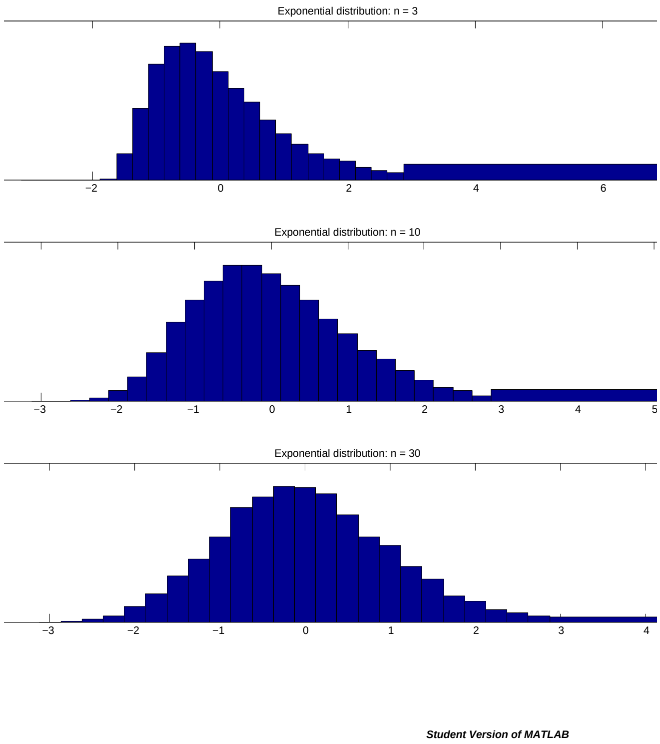


FIGURE 3. Standarized Mean - Exponential Distribution

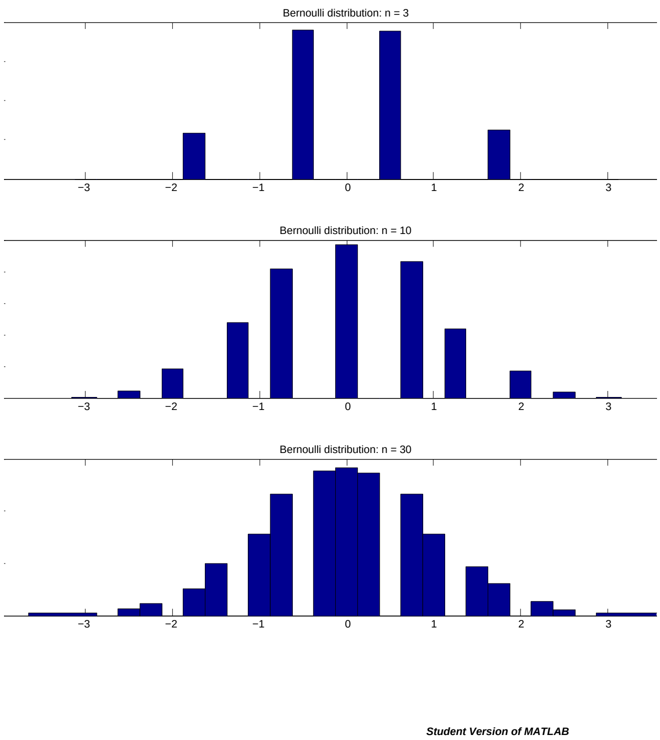


FIGURE 4. Standarized Mean - Bernoulli Distribution

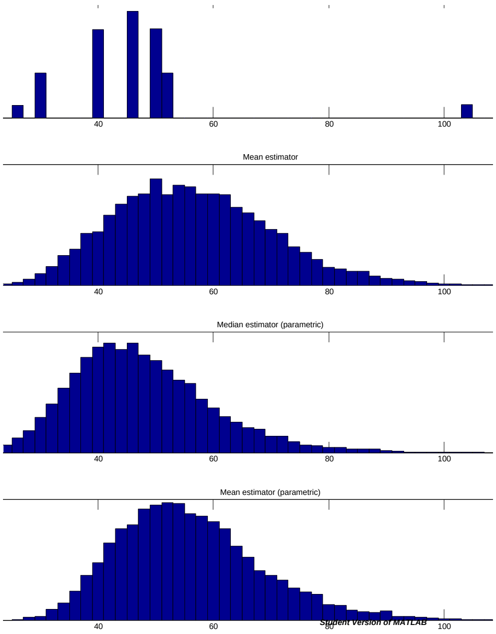


FIGURE 5. Histograms of the two estimators