Solution for the exam in Econometrics A
B.Sc. in Economics 2010-II-R

Academic aim:

The aim of the course is to introduce the students to probability theory and statistics. The aim is for the student to be able to:

- understand the most important basic concepts of probability theory such as: probability, simultaneous-, marginal- and conditional probabilities, distribution, density function, independence, means, variance and covariance and apply these ideas on specific problems.
- know the result from the central limit theory.
- know and recognize the most commonly applied discrete and continuous distributions such as: Bernoulli, binomial, Poisson, multinomial, negative binomial, hypergeometric, geometric, uniform, normal, Chi-squared, exponential, gamma, t-, F-distribution and work with these distributions in relation to specific problems.
- understand the most important statistical concepts such as: random sampling, likelihood function, sufficient statistics, the properties and distributions of statistics, estimation, and maximum likelihood estimation and moment estimation, consistency, confidence interval, hypotheses, test statistics, test probability, level of significance, type I and II errors, power functions.
- perform a simple statistical analysis involving estimation, inference and hypothesis test e.g. the comparison of the means in two populations or test of independence for discrete stochastic variable.
- describe the result of his or her own analysis and considerations in a clear and distinct manner

In order for the student to obtain the highest grade possible, the student must demonstrate the mastery of the above-mentioned skills.

1 Question 1

1. $E(Y_A) = 1.02^5 \cdot 10,000 = 11041$ DKK
   $Var(Y_A) = 0$ since it is a risk free asset.

2. $$E(Y_B) = \left(1.03^5 + 0.03\sqrt{t} \cdot E(Z)\right) \cdot 10,000 = 1.03^5 \cdot 10,000 = 11593$$

$$\begin{aligned}
Var(Y_B) &= Var\left(\left(1.03^5 + 0.03\sqrt{5} \cdot Z\right) \cdot 10,000\right) \\
&= Var\left(10,000 \cdot 1.03^t\right) + Var\left(300\sqrt{5} \cdot Z\right) \\
&= 90000 \cdot 5 \\
&= 450000
\end{aligned}$$

3. What is the probability that asset B gives a lower return than asset A?

$$\Phi\left(\frac{11041 - 11593}{\sqrt{450000}}\right) \approx \Phi(-0.823) \approx 0.2061$$

4.
   (a) The waiting time to market launch is geometric distributed.
   (b) The expected waiting time is $E(T) = \frac{1}{0.1} = 10$ years.

5. Since the investment horizon is 5 years we need to calculate the probability that the biotech product is launched within 5 years

$$\Pr(T \leq 5) = \left(0.9^0 + 0.9^1 + 0.9^2 + 0.9^3 + 0.9^4\right) \cdot 0.1 = 0.40951$$

The expected value of asset B is given by

$$\begin{aligned}
E(Y_B) &= (\Pr(T \leq 5) \cdot 1.30 + \Pr(T > 5) \cdot 1) \cdot 10,000 \\
&= (0.40951 \cdot 1.30 + (1 - 0.40951)) \cdot 10,000 \\
&= 11,229
\end{aligned}$$

Based on expected values, the investor should choose to invest in asset B.

2 Question 2

1. $\Pr(X \geq 10) = 1 - \Pr(X \leq 9) = 1 - 0.070 = 0.93$.
2. It is Erlang distributed (or gamma distributed).
3. $E(T_{10}) = \frac{10}{2} = 5$ hours.
4. We use Bayes theorem

$$\begin{aligned}
\Pr(\text{type=low} \,|X = 14) &= \frac{\Pr(X = 14|\text{ type=low}) \Pr(\text{type=low})}{\Pr(X = 14|\text{ type=low}) \Pr(\text{type=low}) + \Pr(X = 14|\text{ type=high}) \Pr(\text{type=high})} \\
&= \frac{\frac{(1.8 \cdot 7.5)^{14}}{14!} e^{-(1.8 \cdot 7.5)} \cdot \frac{2}{3}}{\frac{(1.8 \cdot 7.5)^{14}}{14!} e^{-(1.8 \cdot 7.5)} \cdot \frac{2}{3} + \frac{(2.4 \cdot 7.5)^{14}}{14!} e^{-(2.4 \cdot 7.5)} \cdot \frac{1}{3}} \\
&= \frac{0.0700160558}{0.0700160558 + 0.0218265347} \\
&= \frac{0.0700160558}{0.0700160558 + 0.0218265347} \\
&= 0.76235
\end{aligned}$$

5. We use Bayes theorem

$$\begin{aligned}
\Pr(\text{type=low } | X=28) &= \frac{\Pr(X=28|\text{ type=low})\Pr(\text{type=low})}{\Pr(X=28|\text{ type=low})\Pr(\text{type=low})+\Pr(X=28|\text{ type=high})\Pr(\text{type=high})} \\
&= \frac{\frac{(1.8\cdot 15)^{28}}{28!}e^{-(1.8\cdot 15)}\cdot\frac{2}{3}}{\frac{(1.8\cdot 15)^{28}}{28!}e^{-(1.8\cdot 15)}\cdot\frac{2}{3}+\frac{(2.4\cdot 15)^{28}}{28!}e^{-(2.4\cdot 15)}\cdot\frac{1}{3}} \\
&= 0.83727
\end{aligned}$$

The probability that it is low effort type increases as more time implies less variance. Hence, it is everything being equal a better idea to evaluate after two days than after one day.

3 Question 3

1. $\hat{p} = \bar{z}_n = \frac{1}{n}\sum_{i=1}^{n} z_i$

   It is a consistent estimator since since $E(\bar{z}_n) \to p$ and $Var(\bar{z}_n) = \frac{p(1-p)}{n} \to 0$ as $n \to \infty$. A confidence interval is an interval around the sample mean where we are confident that the population mean lies. When we have a large sample we can use the desired percentile of a standard normal distribution to form the confidence bounds.

2. $\hat{p} = \frac{55}{121} = 0.4545$

   Since we have a large sample we can write the confidence intervals as below.

$$\hat{p} \pm k \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = 0.4545 \pm 1.96 \cdot \sqrt{\frac{0.4545\cdot(1-0.4545)}{121}} = \begin{cases} 0.3658 \\ 0.5432 \end{cases}$$

   The confidence interval is therefore $[0.3658; 0.5432]$.

3. Since we have a large sample we can use the Z-test. The test statistic is

$$Z = \frac{\hat{p}-0.5}{\sqrt{(1-p_0)\,p_0/n}} = \frac{0.4545-0.5}{\sqrt{(1-0.5)\,0.5/121}} = -1.001$$

   The p value is $2\cdot\Phi(-1.00) = 2\cdot 0.1587 = 0.3174$. Hence, we cannot reject that the gender ratio is 50 per cent. This is in accordance with our confidence interval since 0.5 is clearly within the interval $[0.3658; 0.5432]$.

4. It is with the information provided difficult to assess whether the normal distribution is in agreement with the data. The independence assumption seems reasonable although if some students fail the 2000 exam and retake the exam in 2001 the independence assumption is not met. In addition it might violate the independence assumption if there are more than one teacher for each wave and that one or more teacher teaches both in 2000 and 2001.

5. $H_0 : \sigma^2_{ba} = \sigma^2_{ga}$

   $H_A : \sigma^2_{ba} \neq \sigma^2_{ga}$

$$F = \frac{S^2_{ba}}{S^2_{ga}} = \frac{12.38^2}{9.72^2} = 1.6222$$

   which is lower than $F_{0.995}(65, 54) = 1,99$. Hence, we cannot reject that the variances are the same.

6. The pooled variance is

$$\begin{aligned}
S^2_p &= \frac{(n_1-1)\cdot S^2_{ba}+(n_2-1)\cdot S^2_{ga}}{n_{ba}+n_{ga}-2} \\
&= \frac{(55-1)\cdot 12.38^2+(66-1)\cdot 9.72^2}{55+66-2} \\
&= 121.15
\end{aligned}$$

7. We use the large sample Z-test since $n = 66 + 55 = 121$

$$\begin{aligned} Z &= \frac{\bar{X}_{ga} - \bar{X}_{gb}}{\sqrt{\frac{S_{ga}^2}{n_{ga}} + \frac{S_{gb}^2}{n_{gb}}}} \\ &= \frac{80.15 - 69.94}{\sqrt{\frac{9.72^2}{66} + \frac{16.46^2}{58}}} \\ &= 4.1330 \end{aligned}$$

which implies a p-value of $2 \cdot (1 - \Phi(4.13)) = 0$. Even on a 1 per cent significance level we find that a positive treatment effect for girls.

8. The asymptotic distribution of $\hat{\delta}_b$ and $\hat{\delta}$ are normal since we have a large sample. Due to independence we have

$$\begin{aligned} \hat{\delta}_b &\sim N\left(71.16 - 66.73, \frac{9.92^2}{48} + \frac{12.38^2}{55}\right) \Leftrightarrow \\ \hat{\delta}_b &\sim N(4.43, 4.8368) \end{aligned}$$

$$\begin{aligned} \hat{\delta}_g &\sim N\left(80.15 - 69.94, \frac{16.46^2}{58} + \frac{9.72^2}{66}\right) \Leftrightarrow \\ \hat{\delta}_g &\sim N(10.21, 6.1027) \end{aligned}$$

$$\begin{aligned} \hat{\delta} &\sim N(10.21 - 4.43, 6.1027 + 4.8368) \Leftrightarrow \\ \hat{\delta} &\sim N(5.78, 10.94) \end{aligned}$$

9. Since the asymptotic distribution is normal we can use the standard normal percentiles. The 90% confidence interval becomes

$$5.78 \pm 1.68 \cdot \sqrt{10.94} = \begin{cases} 0.223 \\ 11.337 \end{cases}$$

such that the confidence interval is $[0.223; 11.337]$. Hence, not only do the girls perform better than the boys before and after the experiment in terms of averages, they also improve more on a 10 per cent level.

10. Type I error is rejecting H_0 when H_0 is true whereas Type II error is accepting H_0 when H_0 is false. Since the test statistic is a random variable, there is a chance that the test statistic falls in R (rejection region) or R^C both in the case where the null or the alternative is true. For a given test and a given test statistic we cannot reduce both the probability of Type I and Type II error. Often we prefer to control the size of type 1 error e.g. to 0.05. This will determine the rejection region. For a test with rejection region R, the size of the Type II error for a specific θ in H_A is $\beta = \Pr(\text{accept } H_0 | \theta) = \Pr(R^C | \theta)$ whereas $1 - \beta$ is called the power of the test against the alternative H_A.

    In question 5 we accepted the hypothesis of equal variances. Hence we might be right or we might have made a Type II error. It is not possible to compute the probability of making a Type II error without further assumptions, but if the test is powerful the probability will be low.

    In question 7 we rejected the hypothesis of treatment effect for the girls. Hence we might be right or we might have committed a Type I error. When testing at a 5% level the probability of making a Type I error is 5%.