ISQS 5347 Final Exam

Closed book and notes, no electronic devices. Points (out of 100) in parentheses.

1.(15) Suppose that you have a pile of employee satisfaction surveys coming from five different companies. Each employee answered one of the numbers 1,2,3,4, or 5. The following table shows the distribution of employee satisfaction (on a standard 5-point Likert scale) for five different companies.

Satisfaction	Company Name				
Rating	BankTen	DoggyTreats	CraftyCrafts	AgBus	InternetSavvy
1	20%	30%	20%	20%	20%
2	20%	20%	30%	20%	20%
3	30%	0%	10%	30%	0%
4	20%	0%	10%	20%	0%
5	10%	50%	30%	10%	60%
Total	100%	100%	100%	100%	100%

Now suppose that a survey got separated from the pile, so you do not know what company it was from. What is the probability distribution of the company that it is from, given that you see that the person answered a “4”? Assume a uniform prior. Show how you use Bayesian methods in your answer; in particular, you must identify the likelihood function and the prior distribution function.

Solution: Here the θ is the unknown company – θ is either BankTen, DoggyTreats, CraftyCrafts, AgBus, or InternetSavvy. Since the survey outcome “4” is observed, the likelihood function is as follows:

θ	L(θ|y=4)
BankTen	.2
DoggyTreats	0
CraftyCrafts	.1
AgBus	.2
InternetSavvy	0

Since the prior is uniform, it is as follows:

θ	h(θ)
BankTen	.2
DoggyTreats	.2
CraftyCrafts	.2
AgBus	.2
InternetSavvy	.2

The posterior distribution is proportional to likelihood times prior.

```
θ                             h(θ )x L(θ | y=4)
BankTen                         .2x.2 = .04
DoggyTreats                     .2x 0 = 0
CraftyCrafts                    .2x.1 = .02
AgBus                           .2x.2 = .04
InternetSavvy                   .2x 0 = 0
```

Dividing all by .1 to make the probabilities add to 1.0, we get the posterior distribution:

```
θ                           p(θ | y=4)
BankTen                        .4
DoggyTreats                     0
CraftyCrafts                   .2
AgBus                          .4
InternetSavvy                   0
```

2.(10) What is the result from the "proc print" of the following program? Show your best guess of how the output might look.

```
/* Likert scale probabilities */
%let p1        = .1;
%let p2        = .1;
%let p3        = .2;
%let p4        = .5;
%let p5        = .1;

data simu;
   do i = 1 to 9;
      Y = rand('table', &p1 , &p2, &p3, &p4, &p5);
      output;
    end;
run;

proc print;
run;
```

Solution:

The data that are produced are from a discrete distribution with values 1,2,3,4,5, and probabilities .1,.1.2,.5, .1 respectively. Nine observations will be produced.

So the data might look like this:

4
4
5
3
1
4
4
4
1

3.(15) You will demonstrate that the sample average, $\bar{X}$, is an unbiased estimator of μ, using a simulation study. Assume that $\bar{X}$ is the average of n=10 observations sampled from a N(10,1) distribution. Provide specific details of how you will design this simulation study, and describe what result of the study would demonstrate unbiasedness of $\bar{X}$. You don't have to write SAS code, but it should be very clear from your details precisely what the code will do.

Solution: If $\bar{X}$ is an unbiased estimator of μ, then E($\bar{X}$)=μ. In other words, the mean of the distribution of the possible $\bar{X}$'s is equal to μ. So, we can simulate many $\bar{X}$'s, take the average of them, and check to see whether this average is close to μ. The simulation study will do the following:

1. Sample 10 observations from the N(10,1) distribution, compute the average $\bar{X}$.
2. Sample another 10 observations from the N(10,1) distribution, compute the average $\bar{X}$.

…

1,000,000. Sample another 10 observations from the N(10,1) distribution, compute the average $\bar{X}$.

Now take the average of the 1,000,000 $\bar{X}$'s. If $\bar{X}$ is really an unbiased estimator of μ, then this average should be very, very close to 10. (Note: The law of large numbers states that the sample average is close to the true average with larger sample size, so if the number of simulated samples is increased to, say, one billion, then the average of the $\bar{X}$'s should be even closer to 10.)

4.(10) The likelihood function is $L(\theta)=\theta^2$ and the prior distribution is $h(\theta)=2\theta$, for $0<\theta<1$. Find the posterior distribution for θ.

Solution: The posterior distribution is given by $p(\theta \mid data) \propto L(\theta \mid data)h(\theta)=\theta^2(2\theta)=2\theta^3$, for $0<\theta<1$. To make the area under the curve equal to 1.0, we need to find the area under $2\theta^3$ where $0<\theta<1$. So consider $\int_0^1 2\theta^3 d\theta = 2\frac{\theta^4}{4}\Big|_0^1 = \frac{\theta^4}{2}\Big|_0^1 = \frac{1^4}{2}-\frac{0^4}{2}=\frac{1}{2}$. Dividing $2\theta^3$ by $\frac{1}{2}$ gives the posterior distribution $p(\theta \mid data)=4\theta^3$, for $0<\theta<1$.

5. Suppose E(X)=μ and Var(X)=σ^2. Note that μ and σ^2 are constant (non-random).

5.A.(5) Show E{(X-μ)/σ} = 0. Justify (give the reason for) each step.

Solution:

E{(X-μ)/σ}

= (1/σ) E{(X-μ)} (σ is a constant so this follows from the linearity property of expectation)
= (1/σ) {E(X) - μ} (μ is a constant so this follows from the linearity property of expectation)
= (1/σ) {μ - μ} (by definition of μ)
= 0 (by the rules of algebra).

5.B.(5) Show that Var{(X-μ)/σ} = 1. Justify (give the reason for) each step.

Solution: Var{(X-μ)/σ}

= (1/σ^2) Var{(X-μ)} (σ is a constant so this follows from the linearity property of variance)
= (1/σ^2) {Var(X)} (μ is a constant so this follows from the linearity property of variance)
= (1/σ^2) σ^2 (by definition of σ^2)
= 1 (by the rules of algebra).

6. The following data are from a survey where students were asked whether they would buy a car in the next two years. Here are their answers:

No, No, No, No, Yes, No, No, No, No, No, Yes, No.

In statistical science, one supposes that data are produced by models, but that these models have unknown parameter(s).

6.A.(5) Give a model, with an unknown parameter, that reasonably can be assumed to have produced these data.

Solution: Call the observable data $Y_1, Y_2, \ldots, Y_{12}$. These can be assumed to be an iid sample from the following Bernoulli distribution:

y	p(y)
Yes	θ
No	$1-\theta$

6.B.(5) Describe the meaning of the unknown parameter in terms of the students and car purchasing.

Solution: The parameter θ is the probability that a generic student will answer "Yes". This probability is a characteristic of the student's car-buying intentions, as well as of the process by which students enter the sample. If a large number of students are observed, as produced by this process, then the sample proportion of students answering "Yes" will get closer and closer to this true, unknown process parameter.

An analogy to this process is the repeated flipping of a bent coin, where "Heads" is called "Yes."

7.(10) State two differences between Bayesian statistics and classical statistics (use only one or two sentences for each difference).

Solution: One difference is that prior information about parameters is used in Bayesian statistics, but not in classical statistics.

Another difference is the 95% intervals for θ, say $4.5<\theta<6.7$, can be interpreted as "there is a 95% probability that $4.5<\theta<6.7$" in Bayesian statistics, but you have to say "there is 95% confidence that $4.5<\theta<6.7$" in classical statistics.

Another difference is that you can find the probability that a parameter is greater than 0 using Bayesian statistics, but you can't do that with classical statistics.

Another difference is that it is easy to construct intervals for complicated functions of the parameters like the coefficient of variation σ/μ using Bayesian statistics, but it is much harder to do so using classical statistics.

8. Define the following terms. (4 points each)
8.A. type I error

Solution: A type I error is the act of rejecting the null hypothesis when the null hypothesis is true.

8.B. power

Solution. Power is the probability that you reject the null hypothesis when the null hypothesis is false.

8.C. critical value

Solution: The critical value is a quantile of the null distribution of the test statistic (for example, of a t-distribution) that determines the rejection decision. For example, if the test statistic is greater than the critical value, then you reject the null hypothesis.

8.D. p-value

Solution: The p-value is the probability that you will observe a test statistic as extreme as the value of the test statistic that was actually observed, assuming the null hypothesis is true.

8.E. robustness

Solution: Robustness refers to the strength of a statistical procedure when the assumptions are violated. For example, a two-sample testing procedure is robust to nonnormality if the true significance level is near .05, and if the procedure is reasonably powerful, even when the two distributions are nonnormal.