Exam 1

May 20, 2014

1. Suppose that claim amounts are distributed with mean 20,000 and variance 50,000,000. This year, a total of 100 claims resulted in an aggregate claim amount of 2,250,000. Give the credibility-weighted estimate of mean claim severity next year with a margin of error of 5% with 90% confidence.

   A. 20,000

   B. 21,000

   C. 22,000

   D. 22,150

   E. 22,250

2. The random variable X follows a beta distribution with $a = 1$, $\theta = 1$ and unknown b. Consider X as an estimator for $1/b$. What is the mean-square error?

A.
$$\left(\frac{1}{b(b+1)}\right)^2 - \frac{2}{(b+1)(b+2)} + \left(\frac{1}{b+1}\right)^2$$

B.
$$\left(\frac{1}{b(b+1)}\right)^2 + \frac{2}{(b+1)(b+2)} - \left(\frac{1}{b+1}\right)^2$$

C.
$$\frac{1}{b(b+1)} - \frac{2}{(b+1)(b+2)} + \left(\frac{1}{b+1}\right)^2$$

D.
$$\left(\frac{1}{b(b+1)}\right)^2$$

E.
$$\frac{2}{(b+1)(b+2)} - \left(\frac{1}{b+1}\right)^2$$

3. We have the following data from single Pareto with $\alpha = 3$.

$$20, 35, 70, 90$$

   Using the method of moments, calculate $\hat{\theta}$. Using this value, what is $E[(X-30)_+]$?

   A. Less than 10

   B. Between 10 and 15

   C. Between 15 and 20

   D. Between 20 and 25

   E. More than 25

4. You collect the following sample on lifetimes (in days) of a chile pepper plant:

$$80, 90, 95, 100, 100$$

Calculate the kernel density estimate of $F(95)$, using the uniform kernel with bandwidth 7.

A. 0.5

B. 0.52

C. 0.54

D. 0.56

E. 0.58

5. For a random variable X, you are given the following information.

- $E(X) = 3\theta$
- $Var(X) = 5\theta^2 - 1 \qquad (\theta > \frac{1}{\sqrt{5}})$
- $\hat{\theta}(X) = kX$
- $MSE(\theta) = 4bias^2 - k^2$

What is k given that $k > 0.5$?

A. 0.51

B. 0.55

C. 0.59

D. 0.63

E. 0.68

6. The distribution of the number of claims per policy during a two-year period for 20,000 insurance policies is:

Number of Claims per Policy	Number of Policies
0	7000
1	7000
2	4000
3	2000

Fit a binomial model with parameters m and q using the method of moments. (Round m to the nearest whole integer.) What is the value of the log-likelihood given your estimated parameters?

A. -17,504

B. -20,395

C. -25,310

D. -26,275

E. -30,914

7. Suppose that $X_1, \ldots, X_{100}$ are sampled from a distribution with known variance $\sigma^2 = 625$. Given that $\bar{X} = 100$, give the 95% confidence interval for the population mean μ.

   A. $(91.1, 104.9)$
   B. $(95.1, 104.9)$
   C. $(95.1, 106.2)$
   D. $(91.1, 106.2)$
   E. $(95.9, 104.1)$

8. Suppose you observe the following two values in a sample: 2 and 7. Estimate the mean square error of $\bar{X}$ for estimating the population mean.

   A. 1.8

   B. 2.0

   C. 2.3

   D. 3.1

   E. 3.8

9. Suppose you have the following data:

Year	Exposure (e_j)	# losses (n_j)
1	1000	300
2	2000	500
3	1500	700
4	1800	700

Use a chi-square goodness of fit test to check if the Poisson model is a good fit. Which of the following statements is true about our conclusions from the test at each significance level?

A. The Poisson model is not a good fit at $\alpha < 0.005$.

B. The Poisson model is a good fit at $\alpha = 0.01$ but not at $\alpha = 0.005$.

C. The Poisson model is a good fit at $\alpha = 0.05$ but not at $\alpha = 0.01$.

D. The Poisson model is a good fit at $\alpha = 0.1$ but not at $\alpha = 0.05$.

E. None of the above.

10. A company is trying to model absenteeism of its employees and has found that the probability that any employee will miss work on a given day is 0.05. We assume that the company's 100 employees miss work independently (rather than collectively trying to avoid work!). Use the inversion method to simulate the number of people that will miss work in the next three days, using the following values simulated from a uniform(0,1) distribution.

$$0.02, 0.07, 0.16$$

What is the mean number of absent employees in these three days?

A. 0

B. 1

C. 2

D. 3

E. 4

11. Given the claim count data below, test the hypothesis that claim counts follow the distribution given by $P_0(x)$ using a chi-square goodness-of-fit test.

# Claims	$P_0(x)$	# observations
$0 \leq x < 2$	0.5	55
$2 \leq x < 4$	0.3	30
$4 \leq x < 6$	0.2	25

Which of the statements below is true about the p-value of the test?

A. Less than 0.005

B. Between 0.005 and 0.01

C. Between 0.01 and 0.05

D. Between 0.05 and 0.1

E. Greater than 0.1

12. Suppose Luis has a keen sixth sense about whether or not it will rain in the afternoon. On days when he brings an umbrella, there is a 90% chance that it will rain on his afternoon commute home. On days when he does not bring an umbrella, there is only a 2% chance that it will rain in the afternoon.

    Luis brings his umbrella about once every 10 days. Given that it rained this afternoon, what is the probability that Luis brought an umbrella?

    A. Less than 0.6

    B. Between 0.6 and 0.7

    C. Between 0.7 and 0.8

    D. Between 0.8 and 0.9

    E. More than 0.9

13. Annual claim frequencies follow a Poisson distribution with mean λ. The prior distribution of λ has the following probability density function:

$$\pi(\lambda) = \frac{1}{10} e^{-\lambda/10}$$

5 claims were observed for an individual this year. Determine the Bayes estimator of λ.

A. 5.5

B. 6.0

C. 6.5

D. 7.0

E. 7.5

14. Suppose losses X follow a single-parameter Pareto distribution with parameters $\alpha = 2$ and $\theta = 500$. Calculate $E[(X - 1000)_+]$.

A. 250

B. 267

C. 300

D. 450

E. 500

15. Which of the following statements is false about model selection via a score approach?

    A. The better model tends to have a lower Anderson-Darling test statistic.

    B. The better model tends to have a lower Kolmogorov-Smirnov test statistic.

    C. The better model tends to have a lower chi-square goodness-of-fit p-value.

    D. The better model tends to have a higher Anderson-Darling p-value.

    E. The Schwarz-Bayesian Criterion penalizes for model complexity.

16. The number of accidents per month at a busy intersection follows a Poisson distribution. In the most recent two months (which we assume to be independent), the mean number of accidents is 1.

Let Y denote the mean number of accidents for the next two months (also independent). Determine the maximum likelihood estimator of $Pr(Y > 1)$.

A. 0.50

B. 0.55

C. 0.59

D. 0.63

E. 0.67

17. The number of accidents per month at a busy intersection follows a Poisson distribution. In the most recent two months (which we assume to be independent), the mean number of accidents is 1.

    Calculate the observed information for the monthly rate parameter λ, using the maximum likelihood estimator.

    A. 1

    B. 2

    C. 3

    D. 4

    E. 5

18. Let N be the number of observations. Let κ^2 and γ^2 denote the VHM and EPV for a single observation, while VHM_N and EPV_N are the corresponding values for N observations.

    Which of the following statements is not necessarily true about the credibility factor Z?

    A. $Z = \frac{N}{N+k}$

    B. $Z = \frac{N}{N+EPV/VHM}$

    C. $Z = \frac{N^2\kappa}{N^2\kappa^2+\gamma^2}$

    D. $Z = \frac{VHM_N}{\text{total variance}}$

    E. All of the above are always true.

19. For observation i of a survival study:

- d_i is the left truncation point
- x_i is the observed value if not right-censored
- u_i is the observed value if right-censored

You are given

Observation (i)	d_i	x_i	u_i
1	0	1.2	-
2	0	1.3	-
3	0	-	1.3
4	1	2.5	-
5	2	-	2.7

Determine the Kaplan-Meier Product-Limit estimate of $S(2)$.

A. 0.2

B. 0.3

C. 0.4

D. 0.5

E. 0.6

20. From the previous question, compute the Greenwood approximation for $Var(\hat{S}(2)$. (For convenience, the data is repeated below.)

Observation (i)	d_i	x_i	u_i
1	0	1.2	-
2	0	1.3	-
3	0	-	1.3
4	1	2.5	-
5	2	-	2.7

Determine the Kaplan-Meier Product-Limit estimate of $S(2)$.

A. 0.0001

B. 0.0035

C. 0.0048

D. 0.0079

E. 0.0625

21. A random variable X has the survival function given by

$$S(x) = \frac{\theta^2}{(\theta + 3x)^2} \quad (\theta > 0)$$

You observe the following values for X:

$$1, 2, 2$$

Calculate the maximum likelihood estimate of θ.

A. Less than 8

B. Between 8 and 8.5

C. Between 8.5 and 9

D. Between 9 and 9.5

E. Greater than 9.5

22. Suppose the number claims, N, for a selected risk follows a binomial distribution with $m = 5$ and $q = 0.1$. Claim amount X is distributed with mean 200 and variance 1050. What is the process variance of the pure premium?

A. 16,775

B. 17,525

C. 18,525

D. 18,775

E. 19,035

23. An insurance company sells two types of policies:

Type	A priori proportion	Poisson Annual Claim Frequency
A	p	$\lambda = 0.7$
B	$1 - p$	$\lambda = 1.3$

A randomly selected policyholder has 2 claims in Year 1. What is the Buhlmann credibility factor Z for the same policyholder in Year 2?

A. $\frac{2}{2+\frac{1.3+0.6p}{-0.36p^2+0.36p}}$

B. $\frac{2}{2+\frac{1.3-0.6p}{-0.36p^2+0.36p}}$

C. $\frac{1}{1+\frac{1.3-0.6p}{-0.36p^2+0.36p}}$

D. $\frac{1}{1+\frac{1.3+0.6p}{-0.36p^2+0.36p}}$

E. None of the above.

24. Dr. Onco claims to own a dog that is exceptionally good at identifying patients with cancer. Upon a brief ten-minute meeting with the patient, his dog barks a number of times.

# Barks	Probability of being cancer free
0	0.99
1	0.7
2	0.3
3	0.1

Suppose that Dr. Onco's the number of barks follows the following distribution.

# Barks	A priori probability
0	0.5
1	0.2
2	0.2
3	0.1

What is the proportion of all Dr. Onco's patients that have cancer?

A. Below 0.2

B. Between 0.2 and 0.3

C. Between 0.3 and 0.4

D. Between 0.4 and 0.5

E. More than 0.5

25. The number of claims follows a Poisson distribution with mean λ. The severity of claims follows a gamma distribution with parameters $\alpha = 3$ and unknown θ. Claim count and claim severity are independent. Given that $\pi(\lambda)$ is exponential with mean 2, and $\pi(\theta)$ is Poisson with mean 5, use Buhlmann's credibility for aggregate losses to determine k.

A. 0.57

B. 0.67

C. 1.5

D. 1.75

E. 2

26. Suppose that $X_1, \ldots, X_n$ are iid from uniform$(0, \theta)$. We estimate θ with $\max X_i$. What is the asymptotic variance of this estimator?

A. 0.00

B. 0.27

C. 0.59

D. 0.93

E. 1.25

27. An auto-insurance policy has experienced the following claim severities in its most recent 10 claims.

$$1500, 1500, 2000, 2500, 3000, 3000, 3000, 4000, 4500, \quad 10,000$$

Fit a exponential model using the method of maximum likelihood. What is the estimate of the exponential parameter θ?

A. 2000

B. 2500

C. 3000

D. 3500

E. 4000

28. The distribution of 100 randomly selected policies is given by the following

Number of Claims	Number of Policies
0	43
1	32
2	12
3	8
4	5

Which of the following models best represents this data?

A. Discrete uniform

B. Binomial

C. Poisson

D. Negative binomial

E. Hypergeometric

29. Renters insurance claim amounts in a region follow a Weibull distribution given by

$$F(x) = 1 - e^{-\left(\frac{x}{\theta}\right)^{0.7}} \qquad (x > 0)$$

The following claim amounts are observed (the last one is censored):

$$100, 200, 250, 700, 1000^{+}$$

Find the maximum likelihood estimate of θ.

A. Less than 500

B. Between 500 and 600

C. Between 600 and 700

D. Between 700 and 800

E. More than 800

30. For a particular insurance policy, we have

# claims	0	1	2
Probability	$p/2$	p^2	$1 - p/2 - p^2$

You are given that p follows a beta distribution with $a = 1$, $b = 2$, and $\theta = 1$. Suppose that $X = 1$ claim was observed this year. Calculate the Bayesian credibility estimated number of claims next year.

A. 0

B. 0.5

C. 1

D. 1.5

E. 2

31. Sweet-tooth the Cat likes to sneakily devour a number of jellybeans each day. The following number of jellybeans were eaten each day over the past 10 days:

$$2, 2, 3, 3, 3, 5, 7, 9, 10, 10$$

Let $F_{10}(x)$ denote the ogive of this dataset. What is $F_{10}(6)$?

A. 0.60

B. 0.63

C. 0.65

D. 0.70

E. 0.75

32. Suppose that S_1, S_2, and S_3 are independent and have compound Poisson distributions, where $\lambda_1 = 3$, $\lambda_2 = 4$, and $\lambda_3 = 5$, with underlying severity distributions that are Weibull with $\theta_1 = 1000$, $\theta_2 = 4000$, and $\theta_3 = 5000$ respectively, and $\tau_1 = \tau_2 = \tau_3 = 2$. Find the CDF $F(3000)$ of $S = S_1 + S_2 + S_3$.

A. 0.30

B. 0.43

C. 0.52

D. 0.64

E. 1

33. Consider the following data:

$$20, 37, 92, 102, 130$$

Using a Kolmogorov-Smirnov test, determine whether or not the Weibull($\tau = 1.9, \theta = 86$) distribution is a good fit. Critical values:

α-level	Critical value
0.1	$1.22/\sqrt{n}$
0.05	$1.36/\sqrt{n}$
0.01	$1.63/\sqrt{n}$

Which of the following statements is true for a significance level of 0.05?

A. Our test statistic is below the critical value of 0.61; the Weibull(1.9, 86) distribution is a good fit.

B. Our test statistic is below the critical value of 0.61; the Weibull(1.9, 86) distribution is a poor fit.

C. Our test statistic is above the critical value of 0.61; the Weibull(1.9, 86) distribution is a good fit.

D. Our test statistic is below the critical value of 0.55; the Weibull(1.9, 86) distribution is a good fit.

E. Our test statistic is above the critical value of 0.55; the Weibull(1.9, 86) distribution is a poor fit.

34. The interval (0.23, 0.65) is a 95% log-transformed confidence interval for the cumulate hazard rate function at time t, where the cumulative hazard rate function is estimated using the Nelson-Aalen estimator.

    Find the Nelson-Aalen estimate of $S(t)$.

    A. 0.86

    B. 0.88

    C. 0.90

    D. 0.92

    E. 0.94

35. The probability generating function of a random variable X is given by

$$P_X(z) = \left(\frac{0.8}{1 - 0.2z} \right)^3$$

Find $Pr(X = 2)$.

A. 0.0307

B. 0.0425

C. 0.0573

D. 0.0927

E. 0.1229

36. Given that losses are distributed uniformly on $(0, \theta)$, the following payments occurred:

$$60, 70, 70, 100, 150, 150$$

Consider the cost per loss for a franchise deductible of 50. Estimate θ by matching the average sample payment to the expected payment per loss.

A. 150

B. 163

C. 179

D. 212

E. 257

37. We would like to estimate $F(25)$ correctly within 2 percent with at least 90 percent confidence. Let n be the number of observations needed and let p be the number out of n less than 25.

Which of the following values are possible for n and p?

A. $n = 2000$ and $p = 1100$

B. $n = 2500$ and $p = 1725$

C. $n = 2800$ and $p = 2000$

D. $n = 3000$ and $p = 2000$

E. $n = 3200$ and $p = 2100$

38. The following data are collected for the time it takes (in years) for startups to become profitable.

$$2.4, 3.6, 5.2, 5.3, 5.7, 6.9, 10.1$$

Compute the empirical skewness coefficient.

A. 0.58

B. 0.63

C. 0.69

D. 0.99

E. 1.11

39. Suppose that the annual rate of inflation is 5% for losses that follow an exponential($\theta = 2000$) distribution with a deductible d. If the expected cost per payment next year is 2100, which of the following is (are) a possible value of d?

A. 200

B. 300

C. 400

D. 500

E. All of the above.

40. The annual number of claims follows a negative binomial distribution with $r = 2$ and unknown parameter β. The prior of β has the following cumulative distribution function:

$$F(x) = 1 - \left(\frac{1}{1+x}\right)^{\alpha}$$

A randomly selected policy had x claims in Year 1. What is the Buhlmann credibility estimate of the number of claims for the same policy in Year 2?

A. $0.5(1 + x)$

B. $0.5(1 - x)$

C. $\frac{0.5}{1-x}$

D. $\frac{0.5}{1+x}$

E. None of the above.