Course 4 Examination Questions And Illustrative Solutions

November 2000

1. You fit an invertible first-order moving average model to a time series. The lag-one sample autocorrelation coefficient is −0.35.

Determine an initial guess for q, the moving average parameter.

(A) 0.2

(B) 0.4

(C) 0.6

(D) 0.8

(E) 1.0

2. The following data have been collected for a large insured:

Year	Number Of Claims	Average Claim Size
1	100	10,000
2	200	12,500

Inflation increases the size of all claims by 10% per year.

A Pareto distribution with parameters $\alpha = 3$ and θ is used to model the claim size distribution.

Estimate θ for Year 3 using the method of moments.

(A) 22,500

(B) 23,333

(C) 24,000

(D) 25,850

(E) 26,400

3. You are given the following for a dental insurer:

(i) Claim counts for individual insureds follow a Poisson distribution.

(ii) Half of the insureds are expected to have 2.0 claims per year.

(iii) The other half of the insureds are expected to have 4.0 claims per year.

A randomly selected insured has made 4 claims in each of the first two policy years.

Determine the Bayesian estimate of this insured's claim count in the next (third) policy year.

(A) 3.2

(B) 3.4

(C) 3.6

(D) 3.8

(E) 4.0

4. You are studying the length of time attorneys are involved in settling bodily injury lawsuits. T represents the number of months from the time an attorney is assigned such a case to the time the case is settled.

Nine cases were observed during the study period, two of which were not settled at the conclusion of the study. For those two cases, the time spent up to the conclusion of the study, 4 months and 6 months, was recorded instead. The observed values of T for the other seven cases are as follows:

1 3 3 5 8 8 9

Estimate $Pr[3 \le T \le 5]$ using the Product-Limit estimator.

(A) 0.13

(B) 0.22

(C) 0.36

(D) 0.40

(E) 0.44

5. You are investigating the relationship between per capita consumption of natural gas and the price of natural gas. You gathered data from 20 cities and constructed the following model:

$Y = a + bX + e$, where
Y is per capita consumption,
X is the price, and
e is a normal random error term.

You have determined:

$$\hat{a} = 138.561$$
$$\hat{b} = -1.104$$
$$\Sigma X_i^2 = 90{,}048$$
$$\Sigma Y_i^2 = 116{,}058$$
$$\Sigma x_i^2 = \Sigma\left(X_i - \overline{X}\right)^2 = 10{,}668$$
$$\Sigma y_i^2 = \Sigma\left(Y_i - \overline{Y}\right)^2 = 20{,}838$$
$$\Sigma \hat{e}_i^2 = \Sigma\left(Y_i - \hat{Y}_i\right)^2 = 7{,}832$$

Determine the shortest 95% confidence interval for b.

(A) $(-2.1, -0.1)$

(B) $(-1.9, -0.3)$

(C) $(-1.7, -0.5)$

(D) $(-1.5, -0.7)$

(E) $(-1.3, -0.9)$

6. You have observed the following claim severities:

11.0 15.2 18.0 21.0 25.8

You fit the following probability density function to the data:

$$f(x) = \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{1}{2x}(x-m)^2\right), \quad x > 0, \quad m > 0$$

Determine the maximum likelihood estimate of m.

(A) Less than 17

(B) At least 17, but less than 18

(C) At least 18, but less than 19

(D) At least 19, but less than 20

(E) At least 20

7. The following information comes from a study of robberies of convenience stores over the course of a year:

(i) X_i is the number of robberies of the i^{th} store, with $i = 1, 2, \ldots, 500$.

(ii) $\sum X_i = 50$

(iii) $\sum X_i^2 = 220$

(iv) The number of robberies of a given store during the year is assumed to be Poisson distributed with an unknown mean that varies by store.

Determine the semiparametric empirical Bayes estimate of the expected number of robberies next year of a store that reported no robberies during the studied year.

(A) Less than 0.02

(B) At least 0.02, but less than 0.04

(C) At least 0.04, but less than 0.06

(D) At least 0.06, but less than 0.08

(E) At least 0.08

8. For a study of the excess mortality of smokers, you are given:

(i) 100 smokers are observed for a period of three years.

(ii) The hazard rate $h(t)$ for smokers is modeled as $h(t) = b(t)q(t)$, where $q(t)$ is the reference hazard rate for nonsmokers, and $b(t)$ is the factor indicating excess mortality.

(iii) The table below summarizes the information about the study:

Time Interval	Value Of $q(t)$ Over The Interval	Number Of Smokers At The Beginning Of The Interval	Number Of Deaths Over The Interval
(0,1]	0.022	100	3
(1,2]	0.024	97	3
(2,3]	0.026	94	4

(iv) Deaths during each interval are assumed to occur at the end of the interval.

Calculate $\hat{B}(3)$, the estimate of the cumulative relative excess mortality at $t = 3$.

(A) Less than 2

(B) At least 2, but less than 3

(C) At least 3, but less than 4

(D) At least 4, but less than 5

(E) At least 5

9. You are using a three-point moving average to forecast values of a time series. The last three recorded values of the time series are as follows:

$$y_{98} = 100$$
$$y_{99} = 99$$
$$y_{100} = 101$$

Determine $\hat{y}_{105} - \hat{y}_{104}$, the difference between the five-step-ahead and the four-step-ahead forecasted values.

(A) –0.22

(B) –0.04

(C) 0.00

(D) 0.04

(E) 0.22

10. You are given:

(i) Sample size = 100

(ii) The negative loglikelihoods associated with five models are:

Model	Number Of Parameters	Negative Loglikelihood
Generalized Pareto	3	219.1
Burr	3	219.2
Pareto	2	221.2
Lognormal	2	221.4
Inverse Exponential	1	224.2

(iii) The form of the penalty function is $r \ln\left(\frac{n}{2\pi}\right)$.

Which of the following is the best model, using the Schwartz Bayesian Criterion?

(A) Generalized Pareto

(B) Burr

(C) Pareto

(D) Lognormal

(E) Inverse Exponential

11. For a risk, you are given:

(i) The number of claims during a single year follows a Bernoulli distribution with mean p.

(ii) The prior distribution for p is uniform on the interval [0,1].

(iii) The claims experience is observed for a number of years.

(iv) The Bayesian premium is calculated as 1/5 based on the observed claims.

Which of the following observed claims data could have yielded this calculation?

(A) 0 claims during 3 years

(B) 0 claims during 4 years

(C) 0 claims during 5 years

(D) 1 claim during 4 years

(E) 1 claim during 5 years

12. You are given the following linear regression results:

t	Actual	Fitted
1	77.0	77.6
2	69.9	70.6
3	73.2	70.9
4	72.7	72.7
5	66.1	67.1

Determine the estimated lag 1 serial correlation coefficient after one iteration of the Cochrane-Orcutt procedure.

(A) –0.3

(B) –0.2

(C) –0.1

(D) 0.0

(E) 0.1

13. A sample of ten observations comes from a parametric family $f(x, y; q_1, q_2)$ with loglikelihood function

$$\ln L(q_1, q_2) = \sum_{i=1}^{10} \ln f(x_i, y_i; q_1, q_2) = -2.5q_1^2 - 3q_1q_2 - q_2^2 + 5q_1 + 2q_2 + k,$$

where k is a constant.

Determine the estimated covariance matrix of the maximum likelihood estimator, $\begin{bmatrix} \hat{q}_1 \\ \hat{q}_2 \end{bmatrix}$.

(A) $\begin{bmatrix} 0.5 & 0.3 \\ 0.3 & 0.2 \end{bmatrix}$

(B) $\begin{bmatrix} 20 & -30 \\ -30 & 50 \end{bmatrix}$

(C) $\begin{bmatrix} 0.2 & 0.3 \\ 0.3 & 0.5 \end{bmatrix}$

(D) $\begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$

(E) $\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$

14. For an insurance portfolio, you are given:

(i) For each individual insured, the number of claims follows a Poisson distribution.

(ii) The mean claim count varies by insured, and the distribution of mean claim counts follows a gamma distribution.

(iii) For a random sample of 1000 insureds, the observed claim counts are as follows:

Number Of Claims, n	0	1	2	3	4	5
Number Of Insureds, f_n	512	307	123	41	11	6

$\sum nf_n = 750 \quad \sum n^2 f_n = 1494$

(iv) Claim sizes follow a Pareto distribution with mean 1500 and variance 6,750,000.

(v) Claim sizes and claim counts are independent.

(vi) The full credibility standard is to be within 5% of the expected aggregate loss 95% of the time.

Determine the minimum number of insureds needed for the aggregate loss to be fully credible.

(A) Less than 8300

(B) At least 8300, but less than 8400

(C) At least 8400, but less than 8500

(D) At least 8500, but less than 8600

(E) At least 8600

15.-16. *Use the following information for questions 15 and 16.*

Survival times are available for four insureds, two from Class A and two from Class B. The two from Class A died at times $t = 1$ and $t = 9$. The two from Class B died at times $t = 2$ and $t = 4$.

15. <u>*For question 15 only, you are also given*</u>:

A proportional hazards model is used to model the difference in hazard rates between Class A (Z=0) and Class B (Z=1). No other covariates are used. The (partial) maximum likelihood estimate of the parameter $\boldsymbol{b}$ is b.

Determine the value of Breslow's estimator of the cumulative hazard rate for Class A at time t = 3.

(A) $\frac{1}{2}$

(B) $\frac{7}{12}$

(C) $\frac{1}{2e^b+2}+\frac{1}{2e^b+1}$

(D) $\frac{1}{2e^b+2}+\frac{e^b}{2e^b+1}$

(E) $\frac{e^b}{2e^b+2}+\frac{e^b}{2e^b+1}$

15.-16. *(Repeated for convenience) Use the following information for questions 15 and 16.*

Survival times are available for four insureds, two from Class A and two from Class B. The two from Class A died at times $t = 1$ and $t = 9$. The two from Class B died at times $t = 2$ and $t = 4$.

16. *For question 16 only, you are also given*:

Nonparametric Empirical Bayes estimation is used to estimate the mean survival time for each class. Unbiased estimators of the expected value of the process variance and the variance of the hypothetical means are used.

Estimate *Z,* the Bühlmann credibility factor.

(A) 0

(B) 2/19

(C) 4/21

(D) 8/25

(E) 1

17. You are given the following information about an ARMA(1,1) model:

$$f_1 = 0.3$$

$$q_1 = 0.4$$

Calculate r_2.

(A) –0.10

(B) –0.03

(C) 0.03

(D) 0.04

(E) 0.10

18. A jewelry store has obtained two separate insurance policies that together provide full coverage.

You are given:

(i) The average ground-up loss is 11,100.

(ii) Policy A has an ordinary deductible of 5,000 with no policy limit.

(iii) Under policy A, the expected amount paid per loss is 6,500.

(iv) Under policy A, the expected amount paid per payment is 10,000.

(v) Policy B has no deductible and a policy limit of 5,000.

Given that a loss has occurred, determine the probability that the payment under policy B is 5,000.

(A) Less than 0.3

(B) At least 0.3, but less than 0.4

(C) At least 0.4, but less than 0.5

(D) At least 0.5, but less than 0.6

(E) At least 0.6

19. For a portfolio of independent risks, you are given:

(i) The risks are divided into two classes, Class A and Class B.

(ii) Equal numbers of risks are in Class A and Class B.

(iii) For each risk, the probability of having exactly 1 claim during the year is 20% and the probability of having 0 claims is 80%.

(iv) All claims for Class A are of size 2.

(v) All claims for Class B are of size c, an unknown but fixed quantity.

One risk is chosen at random, and the total loss for one year for that risk is observed. You wish to estimate the expected loss for that same risk in the following year.

Determine the limit of the Bühlmann credibility factor as c goes to infinity.

(A) 0

(B) 1/9

(C) 4/5

(D) 8/9

(E) 1

20. Fifteen cancer patients were observed from the time of diagnosis until the earlier of death or 36 months from diagnosis. Deaths occurred during the study as follows:

Time In Months Since Diagnosis	Number Of Deaths
15	2
20	3
24	2
30	d
34	2
36	1

The Nelson-Aalen estimate $\tilde{H}(35)$ is 1.5641.

Calculate the Aalen estimate of the variance of $\tilde{H}(35)$.

(A) Less than 0.10

(B) At least 0.10, but less than 0.15

(C) At least 0.15, but less than 0.20

(D) At least 0.20, but less than 0.25

(E) At least 0.25

21. You are given the following two regression models, each based on a different population of data:

Model A: $Y_i = A_1 + A_2 X_{2i} + A_3 X_{3i} + e_i$ where $i = 1, 2, \ldots, 30$

Model B: $Y_j = B_1 + B_2 X_{2j} + B_3 X_{3j} + e_j$ where $j = 1, 2, \ldots, 50$

You assume that the variances of the two models are equal and pool the data into one model:

Model G: $Y_p = G_1 + G_2 X_{2p} + G_3 X_{3p} + e_p$ where $p = 1, 2, \ldots, 80$

You calculate R^2_{model} and the error sum of squares, denoted as ESS_{model}, for all three models.

Which of the following is the F statistic for testing the hypothesis that Model A is identical to Model B?

(A) $$F_{3,74} = \frac{(ESS_G - ESS_A - ESS_B)/3}{(ESS_A + ESS_B)/74}$$

(B) $$F_{6,77} = \frac{(ESS_G - ESS_A - ESS_B)/6}{(ESS_A + ESS_B)/77}$$

(C) $$F_{6,74} = \frac{(ESS_G - ESS_A - ESS_B)/6}{(ESS_A + ESS_B)/74}$$

(D) $$F_{3,74} = \frac{(R_G^2 - R_A^2 - R_B^2)/3}{(R_A^2 + R_B^2)/74}$$

(E) $$F_{6,77} = \frac{(R_G^2 - R_A^2 - R_B^2)/6}{(R_A^2 + R_B^2)/77}$$

22. You are given the following information about a random sample:

(i) The sample size equals five.

(ii) The sample is from a Weibull distribution with $t = 2$.

(iii) Two of the sample observations are known to exceed 50, and the remaining three observations are 20, 30 and 45.

Calculate the maximum likelihood estimate of q.

(A) Less than 40

(B) At least 40, but less than 45

(C) At least 45, but less than 50

(D) At least 50, but less than 55

(E) At least 55

23. You are given:

(i) The parameter Λ has an inverse gamma distribution with probability density function:

$$g(l) = 500\, l^{-4} e^{-10/l}, \quad l > 0$$

(ii) The size of a claim has an exponential distribution with probability density function:

$$f(x|\Lambda = l) = l^{-1} e^{-x/l}, \; x > 0, \quad l > 0$$

For a single insured, two claims were observed that totaled 50.

Determine the expected value of the next claim from the same insured.

(A) 5

(B) 12

(C) 15

(D) 20

(E) 25

24. The Product-Limit estimator was used to estimate the survival function for a set of lifetime data:

t_i	d_i	Y_i	$\hat{S}(t)$	$s_S^2(t)$
1	1	20	0.950	0.0026
3	1	19	0.900	0.0056
7	1	18	0.850	0.0088
8	1	15	0.793	0.0136
9	1	14	0.737	0.0191
11	1	12	0.675	0.0267
15	1	6	0.563	0.0600

Confidence Coefficients $c_{.05}(a_L, a_U)$ For 95%
Equal Probability Confidence Bands

		a_L	
a_U	0.08	0.10	0.12
0.40	2.8055	2.7666	2.7309
0.42	2.8178	2.7801	2.7456
0.44	2.8295	2.7931	2.7597
0.46	2.8408	2.8055	2.7732
0.48	2.8517	2.8174	2.7862
0.50	2.8623	2.8290	2.7987
0.52	2.8726	2.8402	2.8108
0.54	2.8826	2.8511	2.8226
0.56	2.8924	2.8618	2.8341

Determine the confidence interval for $S(5)$ within the 95% equal probability linear confidence band for $S(t)$ over the range $3 \le t \le 15$.

(A) (0.62, 1.00)

(B) (0.69, 1.00)

(C) (0.71, 1.00)

(D) (0.73, 1.00)

(E) (0.77, 1.00)

25. You are given the following information about an AR(2) model:

$$r_1 = 0.5$$

$$r_2 = 0.4$$

Determine f_1.

(A) 0.1

(B) 0.2

(C) 0.3

(D) 0.4

(E) 0.5

26. You are given a random sample of two values from a distribution function F:

1 3

You estimate $q(F) = \text{Var}(X)$ using the estimator $g(X_1, X_2) = \sum_{i=1}^{2} (X_i - \overline{X})^2$, where $\overline{X} = \dfrac{X_1 + X_2}{2}$.

Determine the bootstrap approximation to the mean square error.

(A) 0.0

(B) 0.5

(C) 1.0

(D) 2.0

(E) 2.5

27. You are given the following information on towing losses for two classes of insureds, adults and youths:

Exposures

Year	Adult	Youth	Total
1996	2000	450	2450
1997	1000	250	1250
1998	1000	175	1175
1999	1000	125	1125
Total	5000	1000	6000

Pure Premium

Year	Adult	Youth	Total
1996	0	15	2.755
1997	5	2	4.400
1998	6	15	7.340
1999	4	1	3.667
Weighted average	3	10	4.167

You are also given that the estimated variance of the hypothetical means is 17.125.

Determine the nonparametric empirical Bayes credibility premium for the youth class, using the method that preserves total losses.

(A) Less than 5

(B) At least 5, but less than 6

(C) At least 6, but less than 7

(D) At least 7, but less than 8

(E) At least 8

28. Prior to observing any claims, you believed that claim sizes followed a Pareto distribution with parameters $q = 10$ and $a = 1, 2$ or 3, with each value being equally likely.

You then observe one claim of 20 for a randomly selected risk.

Determine the posterior probability that the next claim for this risk will be greater than 30.

(A) 0.06

(B) 0.11

(C) 0.15

(D) 0.19

(E) 0.25

29. Prior to observing any claims, you believed that the survival function for claim sizes followed a Dirichlet distribution with parameter function:

$$a(t,\infty) = 4\left(\frac{10}{10+t}\right)$$

You then observe one claim of 20 for a randomly selected risk.

Determine the posterior probability that the next claim for this risk will be greater than 30.

(A) 0.20

(B) 0.25

(C) 0.30

(D) 0.35

(E) 0.40

30. You are given the following information about an MA(2) model:

$$m = 1.0$$
$$q_1 = 0.3$$
$$q_2 = 0.2$$
$$s_e^2 = 4.0$$

Determine the width of the forecast confidence interval of radius two standard deviations around a forecast of y_t five time periods ahead.

(A) 8.5

(B) 8.8

(C) 9.8

(D) 17.0

(E) 19.6

31. You are given:

(i) $y_i = b x_i + e_i$

$$\text{Var}(e_i) = \left(\frac{x_i}{2}\right)^2$$

(ii)

i	x_i	y_i
1	1	8
2	2	5
3	3	3
4	4	–4

Determine the weighted least squares estimate of b.

(A) 0.4

(B) 0.9

(C) 1.4

(D) 2.0

(E) 2.6

32. You are given the following for a sample of five observations from a bivariate distribution:

(i)

x	y
1	4
2	2
4	3
5	6
6	4

(ii) $\bar{x} = 3.6,\ \bar{y} = 3.8$

A is the covariance of the empirical distribution F_e as defined by these five observations.
B is the maximum possible covariance of an empirical distribution with identical marginal distributions to F_e.

Determine $B - A$.

(A) 0.9

(B) 1.0

(C) 1.1

(D) 1.2

(E) 1.3

33. A car manufacturer is testing the ability of safety devices to limit damages in car accidents.

You are given:

(i) A test car has either front air bags or side air bags (but not both), each type being equally likely.

(ii) The test car will be driven into either a wall or a lake, with each accident type being equally likely.

(iii) The manufacturer randomly selects 1, 2, 3 or 4 crash test dummies to put into a car with front air bags.

(iv) The manufacturer randomly selects 2 or 4 crash test dummies to put into a car with side air bags.

(v) Each crash test dummy in a wall-impact accident suffers damage randomly equal to either 0.5 or 1, with damage to each dummy being independent of damage to the others.

(vi) Each crash test dummy in a lake-impact accident suffers damage randomly equal to either 1 or 2, with damage to each dummy being independent of damage to the others.

One test car is selected at random, and a test accident produces total damage of 1.

Determine the expected value of the total damage for the next test accident, given that the kind of safety device (front or side air bags) and accident type (wall or lake) remain the same.

(A) 2.44

(B) 2.46

(C) 2.52

(D) 2.63

(E) 3.09

34. Phil and Sylvia are competitors in the light bulb business. Sylvia advertises that her light bulbs burn twice as long as Phil's. You were able to test 20 of Phil's bulbs and 10 of Sylvia's. You assumed that the distribution of the lifetime (in hours) of a light bulb is exponential, and separately estimated Phil's parameter as $\hat{q}_P = 1000$ and Sylvia's parameter as $\hat{q}_S = 1500$ using maximum likelihood estimation.

Determine q^*, the maximum likelihood estimate of q_P restricted by Sylvia's claim that $q_S = 2q_P$.

(A) Less than 900

(B) At least 900, but less than 950

(C) At least 950, but less than 1000

(D) At least 1000, but less than 1050

(E) At least 1050

35. You are analyzing a large set of observations from a population.

The true underlying model is:

$$y = 0.1t - z + e$$

You fit a two-variable model to the observations, obtaining:

$$y = 0.3t + e^*$$

You are given:

$$\sum t = 0 \qquad \sum t^2 = 16$$
$$\sum z = 0 \qquad \sum z^2 = 9$$

Estimate the correlation coefficient between z and t.

(A) –0.7

(B) –0.6

(C) –0.5

(D) –0.4

(E) –0.3

36. You study the onset of a disease that includes both left and right censoring, using Turnbull's modification of the Product-Limit estimator.

You are given:

j	Time t_j	Number Left Censored c_j	Number Of Events d_j
4	13	c_4	27
5	14	2	–
6	15	0	–
7	16	2	–
8	>16	0	–

In the first iteration of Turnbull's algorithm, you calculate:

	p_{ij}			
j	i			
	4	5	6	7
1	0.066	0.048	0.038	0.035
2	0.202	0.145	0.116	0.107
3	0.323	0.232	0.184	0.171
4	0.409	0.293	0.233	0.216
5		0.282	0.224	0.208
6			0.205	0.190
7				0.072

$\hat{d}_4 = 30.063$

Determine c_4.

(A) 5

(B) 6

(C) 7

(D) 8

(E) 9

37. Data on 28 home sales yield the fitted model:

$$\hat{Y} = 43.9 + 0.238X_2 - 0.000229X_3 + 0.14718X_4 - 6.68X_5 - 0.269X_6$$

where

Y	=	sales price of home
X_2	=	taxes
X_3	=	size of lot
X_4	=	square feet of living space
X_5	=	number of rooms
X_6	=	age in years

You are given that the estimated variance-covariance matrix (lower-triangular portion) of the variables in the regression is:

	Y	X_2	X_3	X_4	X_5	X_6
Y	20,041.4					
X_2	36,909.0	80,964.2				
X_3	229,662.6	439,511.8	5,923,126.9			
X_4	71,479.2	129,032.9	907,497.1	300,121.4		
X_5	127.2	244.5	1,589.3	532.5	1.3	
X_6	–585.4	–1,420.5	–12,877.4	–1,343.4	0.2	190.9

$\hat{b}_j^*$ is the standardized regression coefficient associated with X_j.

Which of the following is correct?

(A) $\hat{b}_2^* > \hat{b}_3^* > \hat{b}_4^*$

(B) $\hat{b}_2^* > \hat{b}_4^* > \hat{b}_3^*$

(C) $\hat{b}_3^* > \hat{b}_4^* > \hat{b}_2^*$

(D) $\hat{b}_4^* > \hat{b}_2^* > \hat{b}_3^*$

(E) $\hat{b}_4^* > \hat{b}_3^* > \hat{b}_2^*$

38. An insurance company writes a book of business that contains several classes of policyholders.

You are given:

(i) The average claim frequency for a policyholder over the entire book is 0.425.

(ii) The variance of the hypothetical means is 0.370.

(iii) The expected value of the process variance is 1.793.

One class of policyholders is selected at random from the book. Nine policyholders are selected at random from this class and are observed to have produced a total of seven claims. Five additional policyholders are selected at random from the same class.

Determine the Bühlmann credibility estimate for the total number of claims for these five policyholders.

(A) 2.5

(B) 2.8

(C) 3.0

(D) 3.3

(E) 3.9

39. You are given the following information about a study of individual claims:

(i) 20^{th} percentile = 18.25

(ii) 80^{th} percentile = 35.80

Parameters m and s of a lognormal distribution are estimated using percentile matching.

Determine the probability that a claim is greater than 30 using the fitted lognormal distribution.

(A) 0.34

(B) 0.36

(C) 0.38

(D) 0.40

(E) 0.42

40. You are given two random walk models. These models are identical in every respect, except that one includes a known positive drift parameter and the other does not include a drift parameter.

Which of the following statements about these random walk models is incorrect?

(A) For the random walk without drift, all forecasted values from time T are equal.

(B) For the random walk without drift, the standard error of the forecast from time T increases as the forecast horizon increases.

(C) For the random walk with drift, the forecasted values from time T will increase linearly as the forecast horizon increases.

(D) For the random walk with drift, the standard error of a forecasted value from time T is equal to the standard error of the corresponding forecasted value for the random walk without drift.

(E) For the random walk with drift, the standard error of the forecast from time T increases or decreases, depending on the drift parameter, as the forecast horizon increases.

Question #1

Solve $r_1 = -0.35 = -\frac{q}{1+q^2}$ which yields the equation $0.35q^2 - q + 0.35 = 0$.
The two solutions of this quadratic equation are 0.408 and 2.449. To be invertible, the parameter must be less than one in absolute value, so the answer is 0.408.

(B)

Question #2

First inflate the claims. Then there are 100 averaging 10,000 (1.21) = 12,100 for a total of 1,210,000 and 200 averaging 12,500 (1.1) = 13,750 for a total of 2,750,000. The grand total is 3,960,000 on 300 claims for an average of 13,200. The mean of this Pareto distribution is $\frac{q}{2}$.

Setting this equal to 13,200 yields $\hat{q} = 26,400$.

(E)

Question #3

From Bayes' Theorem:

$$\Pr(l = 2|4,4) = \frac{\Pr(4,4|l = 2)\Pr(l = 2)}{\Pr(4,4|l = 2)\Pr(l = 2) + \Pr(4,4|l = 4)\Pr(l = 4)}$$

$$= \frac{\left(e^{-2}2^4 / 4!\right)^2 (0.5)}{\left(e^{-2}2^4 / 4!\right)^2 (0.5) + \left(e^{-4}4^4 / 4!\right)^2 (0.5)}$$

$$= 0.17578$$

Because $E(X) = l$, the expected next claim count is the posterior expected value of l which is 0.17578(2) + 0.82422(4) = 3.648.

(C)

Question #4

The data reveal death times of 1, 3 (2 deaths), 5, 8 (2 deaths), and 9 and withdrawal times of 4 and 6. The Product-Limit estimate produces a discrete distribution with probability at each death time. The probability of dying at time 3 is given by (8/9)(2/8) = 2/9. The probability of dying at time 5 is given by (8/9)(6/8)(1/5) = 2/15. The total probability of dying between times 3 and 5 inclusive is 2/9 + 2/15 = 16/45 = 0.356.

(C)

Question #5

The following facts are needed:

$$s^2 = \frac{7832}{18} = 435.11,\ t_{18,\ .05} = 2.101.$$

The confidence interval is $-1.104 \pm 2.101\sqrt{\frac{435.11}{10{,}668}}$ or -1.104 ± 0.424.

(D)

Question #6

$$L(m) \propto \exp\left[-\frac{1}{22}(11-m)^2 - \frac{1}{30.4}(15.2-m)^2 - \frac{1}{36}(18-m)^2 - \frac{1}{42}(21-m)^2 - \frac{1}{51.6}(25.8-m)^2\right]$$

$$\frac{d}{dm}\ln L(m) = \frac{1}{11}(11-m) + \frac{1}{15.2}(15.2-m) + \frac{1}{18}(18-m) + \frac{1}{21}(21-m) + \frac{1}{25.8}(25.8-m)$$

and setting this equal to zero yields $5 - 0.298633m = 0$ for $\hat{m} = 16.74$.

(A)

Question #7

$\hat{v} = \bar{x} = \dfrac{50}{500} = 0.1$ and

$\hat{a} = s^2 - 0.1 = \dfrac{\left[220 - 500(0.1)^2\right]}{499} - 0.1 = 0.33086.$

Then $\hat{Z} = \dfrac{1}{1 + \dfrac{0.1}{0.33086}} = 0.76791.$

The credibility estimate is then $0.76791(0) + 0.23209(0.1) = 0.023209.$

(B)

Question #8

$Q(1) = 0.022(100) = 2.2$

$Q(2) = 0.024(97) = 2.328$

$Q(3) = 0.026(94) = 2.444$

Then $\hat{B}(3) = \dfrac{3}{2.2} + \dfrac{3}{2.328} + \dfrac{4}{2.444} = 4.289.$

(D)

Question #9

$$\hat{y}_{101} = \frac{100+99+101}{3} = 100,$$
$$\hat{y}_{102} = \frac{99+101+100}{3} = 100,$$
$$\hat{y}_{103} = \frac{101+100+100}{3} = 100.33,$$
$$\hat{y}_{104} = \frac{100+100+100.33}{3} = 100.11,$$
$$\hat{y}_{105} = \frac{100+100.33+100.11}{3} = 100.15.$$

Then $\hat{y}_{105} - \hat{y}_{104} = 0.04.$

(D)

Question #10

The score is $\ln L - r\ln\left(\frac{n}{2p}\right) = \ln L - 2.7673r.$

The scores are:

Generalized Pareto: $-219.1 - 3(2.7673) = -227.40$

Burr: $-219.2 - 3(2.7673) = -227.50$

Pareto: $-221.2 - 2(2.7673) = -226.73$

Lognormal: $-221.4 - 2(2.7673) = -226.93$

Inverse exponential: $-224.2 - 2.7673 = -226.97$

The highest value is for the Pareto.

(C)

Question #11

The posterior distribution is

$p(p|n,x) \propto f(x|n,p)p(p) \propto p^x(1-p)^{n-x}1$, where n is the number of years and x is the number of claims. This is a beta distribution. The Bayesian estimate is the posterior mean, which is $\dfrac{x+1}{n+2}$. The only combination among the five answer choices that produces 0.2 is $x = 0$ and $n = 3$.

(A)

Question #12

The residuals are $-0.6, -0.7, 2.3, 0, -1$.

The estimate of the lag-one serial correlation coefficient is

$$\frac{\sum_{t=2}^{n} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=2}^{n} \hat{e}_{t-1}^2} = \frac{0.42 - 1.61 + 0 + 0}{0.36 + 0.49 + 5.29 + 0} = -0.1938.$$

(B)

Question #13

$$\frac{\partial}{\partial q_1} \ln L = -5q_1 - 3q_2 + 5, \ \frac{\partial}{\partial q_2} \ln L = -3q_1 - 2q_2 + 2, \ \frac{\partial}{\partial q_1^2} \ln L = -5, \ \frac{\partial^2}{\partial q_2^2} \ln L = -2, \ \frac{\partial^2}{\partial q_1 \partial q_2} \ln L = -3.$$

The information matrix is $\begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$ and the covariance matrix is its inverse: $\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$.

(E)

Question #14

$$x = E(X) = E(N)q_Y = 0.75(1500) = 1125.$$

$$s^2 = Var(X) = E(N)s_Y^2 + Var(N)q_Y^2 = 0.75(6{,}750{,}000) + 0.932432(1500)^2 = 7{,}160{,}473.$$

$$n \geq \left(\frac{1.96}{0.05}\right)^2 \frac{7{,}160{,}473}{1125^2} = 8{,}694.$$

(E)

Question #15

$$W(1,b) = 2 + 2e^b,\ W(2,b) = 1 + 2e^b,\ \hat{H}_0(3) = \frac{1}{2e^b + 2} + \frac{1}{2e^b + 1}.$$

(C)

Question #16

$$\hat{v} = \frac{1}{2(2-1)}\left[(1-5)^2 + (9-5)^2 + (2-3)^2 + (4-3)^2\right] = 17.$$

$$\hat{a} = \frac{1}{2-1}\left[(5-4)^2 + (3-4)^2\right] - \frac{17}{2} = -6.5 < 0.$$

When this happens, the convention is to set $\hat{Z} = 0$.

(A)

Question #17

$$r_2 = f_1 \frac{(1 - f_1 q_1)(f_1 - q_1)}{1 + q_1^2 - 2 f_1 q_1} = 0.3 \frac{(1 - 0.12)(0.3 - 0.4)}{1 + 0.16 - 0.24} = -0.029.$$

(B)

Question #18

$$10{,}000 = \frac{6{,}500}{1 - F(5{,}000)}, F(5{,}000) = 0.35, \ \Pr(X \geq 5{,}000) = 0.65.$$

(E)

Question #19

For class A, the mean is $0.2(2) = 0.4$ and the variance is $0.2(2)^2 - 0.16 = 0.64$.

For class B, the mean is $0.2c$ and the variance is $0.2c^2 - 0.04c^2 = 0.16c^2$.

Then, $m = 0.5(0.4) + 0.5(0.2c) = 0.2 + 0.1c$.

Also, $a = 0.5(0.4 - 0.2 - 0.1c)^2 + 0.5(0.2c - 0.2 - 0.1c)^2 = 0.04 - 0.04c + 0.01c^2$ and $v = 0.5(0.64) + 0.5(0.16c^2) = 0.32 + 0.08c^2$.

Then, $$Z = \frac{1}{1 + \dfrac{0.32 + 0.08c^2}{0.04 - 0.04c + 0.01c^2}} = \frac{0.04 - 0.04c + 0.01c^2}{0.36 - 0.04c + 0.09c^2}.$$

As c goes to infinity, Z goes to 1/9.

(B)

Question #20

$$\tilde{H}(35) = \frac{2}{15} + \frac{3}{13} + \frac{2}{10} + \frac{d}{8} + \frac{2}{8-d} = 1.5641.$$

This leads to $\frac{d}{8} + \frac{2}{8-d} = 1$ or $d^2 - 16d + 48 = 0.$

The solution is $d = 4$.

Then the estimated variance is $\frac{2}{15^2} + \frac{3}{13^2} + \frac{2}{10^2} + \frac{4}{8^2} + \frac{2}{4^2} = 0.2341.$

(D)

Question #21

The F statistic, with 3 degrees of freedom in the numerator (the number of restrictions), and 74 degrees of freedom in the denominator (the number of degrees of freedom in the unrestricted regression), is

$$\frac{(ESS_G - ESS_A - ESS_B)/3}{(ESS_A + ESS_B)/74}.$$

(A)

Question #22

$$L(q) = f(20)f(30)f(45)[1 - F(50)]^2 \propto q^{-2}e^{-(20/q)^2}q^{-2}e^{-(30/q)^2}q^{-2}e^{-(45/q)^2}\left[e^{-(50/q)^2}\right]^2.$$

This is equal to $q^{-6}e^{-8325/q^2}$ and the derivative of its logarithm is $-6/q + 16650/q^3$.

Setting it equal to zero yields $\hat{q} = 52.68$.

(D)

Question #23

$g(l|x_1,x_2) \propto l^{-1}e^{-x_1/l}\, l^{-1}e^{-x_2/l}\, l^{-4}e^{-10/l} = l^{-6}e^{-60/l}$.

This is an inverse gamma distribution with $a = 5$ and $q = 60$. The expected value of this random variable is 60/4 = 15. The next claim has an exponential distribution with mean l, which itself has a mean of 15.

(C)

Question #24

$a_L = \frac{20(0.0056)}{1+20(0.0056)} = 0.1007$, $a_U = \frac{20(0.06)}{1+20(0.06)} = 0.545$, $c_{0.05}(0.1, 0.545) = 2.854$.

The lower confidence limit is $0.9 - 2.854\sqrt{0.0056}(0.9) = 0.71$.

The upper limit is 1.09, which is taken as 1 (the maximum possible value).

(C)

Question #25

$r_1 = f_1 + f_2 r_1, r_2 = f_1 r_1 + f_2$.

This yields $0.5 = f_1 + 0.5 f_2$, $0.4 = 0.5 f_1 + f_2$.

The solution is $f_1 = 0.4$.

(D)

Question #26

The variance from the sample is $q(F_e) = 0.5\left[(1-2)^2 + (3-2)^2\right] = 1.$

There are four possible bootstrap samples:

Sample	$g(x_1, x_2)$	MSE
1,1	0	1
1,3	2	1
3,1	2	1
3,3	0	1

And so the average MSE is 1.

(C)

Question #27

$$\hat{v} = \Big[2000(0-3)^2 + 1000(5-3)^2 + 1000(6-3)^2 + 1000(4-3)^2$$
$$+450(15-10)^2 + 250(2-10)^2 + 175(15-10)^2 + 125(1-10)^2\Big]/6 = 12{,}292.$$

Then $\hat{Z}_1 = \dfrac{5000}{5000 + 12292/17.125} = 0.87447$, $\hat{Z}_2 = \dfrac{1000}{1000 + 12292/17.125} = 0.58215.$

Also, $\hat{m} = \dfrac{0.87447(3) + 0.58215(10)}{0.87447 + 0.58215} = 5.79762.$

For the youth class, the estimate is $0.58215(10) + 0.41785(5.79762) = 8.244.$

(E)

Question #28

$f(20|a=1) = 10/30^2 = 0.011111,$

$f(20|a=2) = 2(10)^2/30^3 = 0.007407,$

$f(20|a=3) = 3(10)^3/30^4 = 0.003704.$

The total of these three probabilities is 0.022222 and thus the posterior probabilities are 1/2, 1/3, and 1/6. The following table yields the posterior expectation:

a	Posterior Probability	$\Pr(X > 30\|a)$
1	1/2	$10/40 = 0.25$
2	1/3	$(10/40)^2 = 0.0625$
3	1/6	$(10/40)^3 = 0.015625$

The weighted average is 0.1484.

(C)

Question #29

$$\tilde{S}_D(30) = \frac{a(30,\infty)+0}{a(0,\infty)+1} = \frac{4(10/40)}{4+1} = 0.2$$

(A)

Question #30

The width of the confidence interval is $2(2)\sqrt{4}\sqrt{1+0.3^2+0.2^2} = 8.5.$

(A)

Question #31

The weighted least squares estimate is

$$\frac{\sum (x_i / s_i)(y_i / s_i)}{\sum (x_i / s_i)^2} = \left(\frac{8}{0.25} + \frac{10}{1} + \frac{9}{2.25} - \frac{16}{4}\right)\left(\frac{1}{0.25} + \frac{4}{1} + \frac{9}{2.25} + \frac{16}{4}\right)^{-1} = 2.625.$$

(E)

Question #32

To maximize the covariance, sort the y values from smallest to largest. The pairs are then (1,2), (2,3), (4,4), (5,4), and (6,6).

The expected product is $\frac{(2+6+16+20+36)}{5} = 16.$

The covariance is $16 - 3.6(3.8) = 2.32.$

For the original data, the covariance is $\frac{(4+4+12+30+24)}{5} - 3.6(3.8) = 1.12.$

The difference is 1.2.

(D)

Question #33

Let S = side airbag, F = front airbag, W = wall-impact, L = lake-impact, and the number is the number of crash test dummies. The first table yields the posterior probability of each configuration given that there was total damage of 1.

Configuration	Prob (damage = 1)	Prior prob	Product	Posterior*
SW2	0.25	1/8	1/32	2/7
SW4	0.00	1/8	0	0
SL2	0.00	1/8	0	0
SL4	0.00	1/8	0	0
FW1	0.50	1/16	1/32	2/7
FW2	0.25	1/16	1/64	1/7
FW3	0.00	1/16	0	0
FW4	0.00	1/16	0	0
FL1	0.50	1/16	1/32	2/7
FL2	0.00	1/16	0	0
FL3	0.00	1/16	0	0
FL4	0.00	1/16	0	0

*The posterior probability is the product divided by the sum of the items in the product column, 7/64.

For the next accident, the posterior probabilities are 2/7 for SW, 3/7 for FW, and 2/7 for FL.

For an SW configuration, the expected next damage is $\frac{1.5+3}{2} = 2.25$ (based on the average of the expectations for 2 and 4 dummies).

For FW, it is $\frac{(0.75+1.5+2.25+3)}{4} = 1.875.$

For FL, it is $\frac{(1.5+3+4.5+6)}{4} = 3.75.$

The posterior mean is $(2/7)(2.25)+(3/7)(1.875)+(2/7)(3.75) = 2.52.$

(C)

Question #34

Because the parameter estimate is the sample mean, the sample means are 1000 and 1500 for Phil and Sylvia respectively. The likelihood function is

$$L(q) = \prod_{i=1}^{20} q^{-1} e^{-x_i/q} \prod_{j=1}^{10} (2q)^{-1} e^{-y_j/2q} \infty q^{-30} e^{-20,000/q - 15,000/2q}$$

because the totals for Phil and Sylvia are 20,000 and 15,000 respectively. Setting the derivative of the logarithm of the likelihood equal to zero gives the equation

$$-\frac{30}{q} + \frac{27{,}500}{q^2} = 0,\ q^* = \frac{27{,}500}{30} = 917.$$

(B)

Question #35

$$0.3 = 0.1 - 1\frac{\operatorname{Cov}(t,z)}{\operatorname{Var}(t)},\ -0.2 = \frac{\operatorname{Cov}(t,z)}{16/n},\ \operatorname{Cov}(t,z) = \frac{-3.2}{n}.$$

Also, $\operatorname{Var}(z) = \frac{9}{n}$.

Then, $\operatorname{Corr}(t,z) = \frac{-3.2/n}{\sqrt{(16/n)(9/n)}} = -0.267.$

(E)

Question #36

$$\hat{d}_4 = d_4 + \sum c_i p_{i4},\ 30.063 = 27 + c_4(0.409) + 2(0.293) + 0(0.233) + 2(0.216),\ c_4 = 5.$$

(A)

Question #37

The formula is:

$$\hat{b}_j^* = \frac{\hat{b}_j s_{X_j}}{s_Y},$$

$$\hat{b}_2^* = 0.238\sqrt{\frac{80964.2}{20041.4}} = 0.478,$$

$$\hat{b}_3^* = -0.000229\sqrt{\frac{5923126.9}{20041.4}} = -0.004,$$

$$\hat{b}_4^* = 0.14718\sqrt{\frac{300121.4}{20041.4}} = 0.570.$$

Then $\hat{b}_4^* > \hat{b}_2^* > \hat{b}_3^*$.

(D)

Question #38

$$Z = \frac{9}{9 + 1.793/0.37} = 0.65.$$

The estimate for one policyholder is $0.65(7/9) + 0.35(0.425) = 0.6543$.

For 5 policyholders, it is $5(0.6543) = 3.27$.

(D)

Question #39

The two equations to solve are $-0.842 = \dfrac{\ln 18.25 - m}{s}$ and $0.842 = \dfrac{\ln 35.8 - m}{s}$.

The solutions are $\hat{s} = 0.4$ and $\hat{m} = 3.241$.

Then $\Pr(X > 30) = 1 - \Phi\left(\dfrac{\ln 30 - 3.241}{0.4}\right) = 1 - \Phi(0.4) = 1 - 0.6554 = 0.3446$.

(A)

Question #40

The forecast error always increases with the forecast horizon, so E is false.

(E)

ANSWER KEY

Question Number	Answer	Question Number	Answer
1	B	21	A
2	E	22	D
3	C	23	C
4	C	24	C
5	D	25	D
6	A	26	C
7	B	27	E
8	D	28	C
9	D	29	A
10	C	30	A
11	A	31	E
12	B	32	D
13	E	33	C
14	E	34	B
15	C	35	E
16	A	36	A
17	B	37	D
18	E	38	D
19	B	39	A
20	D	40	E