Course 4 Fall 2004 Society of Actuaries

****BEGINNING OF EXAMINATION****

1. You are given:

(i) The annual number of claims for an insured has probability function:

$$
p(x) = {3 \choose x} q^x (1-q)^{3-x}, \quad x = 0, 1, 2, 3
$$

(ii) The prior density is $\pi(q) = 2q$, $0 < q < 1$.

A randomly chosen insured has zero claims in Year 1.

Using Bühlmann credibility, estimate the number of claims in Year 2 for the selected insured.

- (A) 0.33
- (B) 0.50
- (C) 1.00
- (D) 1.33
- (E) 1.50

2. You are given the following random sample of 13 claim amounts:

99 133 175 216 250 277 651 698 735 745 791 906 947

Determine the smoothed empirical estimate of the $35th$ percentile.

- (A) 219.4
- (B) 231.3
- (C) 234.7
- (D) 246.6
- (E) 256.8

- (i) *Y* is the annual number of discharges from a hospital.
- (ii) *X* is the number of beds in the hospital.
- (iii) Dummy D is 1 if the hospital is private and 0 if the hospital is public.
- (iv) The proposed model for the data is $Y = \beta_1 + \beta_2 X + \beta_3 D + \varepsilon$.
- (v) To correct for heteroscedasticity, the model $Y/X = \beta_1/X + \beta_2 + \beta_3 D/X + \varepsilon/X$ is fitted to *N* = 393 observations, yielding $\hat{\beta}_2 = 3.1$, $\hat{\beta}_1 = -2.8$ and $\hat{\beta}_3 = 28$.
- (vi) For the fit in (v) above, the matrix of estimated variances and covariances of $\hat{\beta}_2$, $\hat{\beta}_1$ and $\hat{\beta}_3$ is:

Determine the upper limit of the symmetric 95% confidence interval for the difference between the mean annual number of discharges from private hospitals with 500 beds and the mean annual number of discharges from public hospitals with 500 beds.

- $(A) 6$
- (B) 31
- (C) 37
- (D) 40
- (E) 67
- **4.** For observation *i* of a survival study:
	- \bullet *d_i* is the left truncation point
	- x_i is the observed value if not right censored
	- \bullet *u_i* is the observed value if right censored

Determine the Kaplan-Meier Product-Limit estimate, S_{10} (1.6).

- (A) Less than 0.55
- (B) At least 0.55, but less than 0.60
- (C) At least 0.60, but less than 0.65
- (D) At least 0.65, but less than 0.70
- (E) At least 0.70

(ii) Class 1 has twice as many claims as Class 2.

A claim of 250 is observed.

Determine the Bayesian estimate of the expected value of a second claim from the same policyholder.

- (A) Less than 10,200
- (B) At least 10,200, but less than 10,400
- (C) At least 10,400, but less than 10,600
- (D) At least 10,600, but less than 10,800
- (E) At least 10,800

6. You are given the following three observations:

$$
\begin{matrix} 0.74 & 0.81 & 0.95 \end{matrix}
$$

You fit a distribution with the following density function to the data:

$$
f(x) = (p+1)x^p, \quad 0 < x < 1, \ p > -1
$$

Determine the maximum likelihood estimate of *p*.

- (A) 4.0
- (B) 4.1
- (C) 4.2
- (D) 4.3
- (E) 4.4
- **7.** Which of the following statements about time series is false?
	- (A) If the characteristics of the underlying stochastic process change over time, then the process is nonstationary.
	- (B) A stationary process is one whose joint distribution and conditional distribution functions are both invariant with respect to displacement in time.
	- (C) If a time series is stationary, then its mean, variance and, for any lag *k*, covariance must also be stationary.
	- (D) If the autocorrelation function for a time series is zero (or close to zero) for all lags $k > 0$, then no model can provide useful minimum mean-square-error forecasts of future values other than the mean.
	- (E) A homogeneous nonstationary time series will always be nonstationary, regardless of how many times it is differenced.
- **8.** You are given the following sample of claim counts:
	- 0 0 1 2 2

You fit a binomial (m, q) model with the following requirements:

- (i) The mean of the fitted model equals the sample mean.
- (ii) The $33rd$ percentile of the fitted model equals the smoothed empirical $33rd$ percentile of the sample.

Determine the smallest estimate of *m* that satisfies these requirements.

- $(A) 2$
- (B) 3
- $(C) 4$
- (D) 5
- (E) 6

9. Members of three classes of insureds can have 0, 1 or 2 claims, with the following probabilities:

A class is chosen at random, and varying numbers of insureds from that class are observed over 2 years, as shown below:

Determine the Bühlmann-Straub credibility estimate of the number of claims in Year 3 for 35 insureds from the same class.

- (A) 10.6
- (B) 10.9
- (C) 11.1
- (D) 11.4
- (E) 11.6

10. You are given the following random sample of 30 auto claims:

You test the hypothesis that auto claims follow a continuous distribution $F(x)$ with the following percentiles:

You group the data using the largest number of groups such that the expected number of claims in each group is at least 5.

Calculate the chi-square goodness-of-fit statistic.

- (A) Less than 7
- (B) At least 7, but less than 10
- (C) At least 10, but less than 13
- (D) At least 13, but less than 16
- (E) At least 16
- **11.** For the model $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$, you are given:
	- (i) $r_{YX_2} = 0.4$
	- (ii) $r_{YX_3} \cdot x_2 = -0.4$

Determine R^2 .

- (A) 0.03
- (B) 0.16
- (C) 0.29
- (D) 0.71
- (E) 0.84

12. The interval (0.357, 0.700) is a 95% log-transformed confidence interval for the cumulative hazard rate function at time *t*, where the cumulative hazard rate function is estimated using the Nelson-Aalen estimator.

Determine the value of the Nelson-Aalen estimate of *S*(*t*).

- (A) 0.50
- (B) 0.53
- (C) 0.56
- (D) 0.59
- (E) 0.61
- **13.** You are given:
	- (i) The number of claims observed in a 1-year period has a Poisson distribution with mean θ .
	- (ii) The prior density is:

$$
\pi(\theta) = \frac{e^{-\theta}}{1 - e^{-k}}, \quad 0 < \theta < k
$$

(iii) The unconditional probability of observing zero claims in 1 year is 0.575.

Determine *k*.

- (A) 1.5
- (B) 1.7
- (C) 1.9
- (D) 2.1
- (E) 2.3

14. The parameters of the inverse Pareto distribution

$$
F(x) = \left[x/(x+\theta)\right]^{\tau}
$$

are to be estimated using the method of moments based on the following data:

15 45 140 250 560 1340

Estimate θ by matching *k*th moments with $k = -1$ and $k = -2$.

- (A) Less than 1
- (B) At least 1, but less than 5
- (C) At least 5, but less than 25
- (D) At least 25, but less than 50
- (E) At least 50

15. You are given the following information about a stationary time-series model:

$$
\rho_1 = -0.310
$$

\n
$$
\rho_2 = -0.155
$$

\n
$$
\rho_k = 0, \quad k = 3, 4, 5, \dots
$$

You are also given that $\theta_1 + \theta_2 = 0.7$.

Determine θ_1 .

- (A) 0.1
- (B) 0.2
- (C) 0.3
- (D) 0.4
- (E) 0.5

16. A sample of claim amounts is {300, 600, 1500}. By applying the deductible to this sample, the loss elimination ratio for a deductible of 100 per claim is estimated to be 0.125.

You are given the following simulations from the sample:

Determine the bootstrap approximation to the mean square error of the estimate.

- (A) 0.003
- (B) 0.010
- (C) 0.021
- (D) 0.054
- (E) 0.081

Determine the nonparametric empirical Bayes credibility factor, *Z*, for Company III.

- (A) Less than 0.2
- (B) At least 0.2, but less than 0.4
- (C) At least 0.4, but less than 0.6
- (D) At least 0.6, but less than 0.8
- (E) At least 0.8

18. Let $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_m$ denote independent random samples of losses from Region 1 and Region 2, respectively. Single-parameter Pareto distributions with $\theta = 1$, but different values of α , are used to model losses in these regions.

Past experience indicates that the expected value of losses in Region 2 is 1.5 times the expected value of losses in Region 1. You intend to calculate the maximum likelihood estimate of α for Region 1, using the data from both regions.

Which of the following equations must be solved?

(A)
$$
\frac{n}{\alpha} - \sum \ln(x_i) = 0
$$

(B)
$$
\frac{n}{\alpha} - \sum \ln(x_i) + \frac{m(\alpha+2)}{3\alpha} - \frac{2\sum \ln(y_i)}{(\alpha+2)^2} = 0
$$

(C)
$$
\frac{n}{\alpha} - \sum \ln(x_i) + \frac{2m}{3\alpha(\alpha+2)} - \frac{2\sum \ln(y_i)}{(\alpha+2)^2} = 0
$$

(D)
$$
\frac{n}{\alpha} - \sum \ln(x_i) + \frac{2m}{\alpha(\alpha + 2)} - \frac{6\sum \ln(y_i)}{(\alpha + 2)^2} = 0
$$

(E)
$$
\frac{n}{\alpha} - \sum \ln(x_i) + \frac{3m}{\alpha(3-\alpha)} - \frac{6\sum \ln(y_i)}{(3-\alpha)^2} = 0
$$

- **19.** You are given the following information about a linear regression model:
	- (i) The unit of measurement is a region, and the number of regions in the study is 37.
	- (ii) The dependent variable is a measure of workers' compensation frequency, while the three independent variables are a measure of employment, a measure of unemployment rate and a dummy variable indicating the presence or absence of vigorous cost-containment efforts.
	- (iii) The model is fitted separately to the group of 18 largest regions and to the group of 19 smallest regions (by population). The ESS resulting from the first fit is 4053, while the ESS resulting from the second fit is 2087.
	- (iv) The model is fitted to all 37 regions, and the resulting ESS is 10,374.

The null hypothesis to be tested is that the pooling of the regions into one group is appropriate.

Which of the following is true?

- (A) The *F* statistic has 4 numerator degrees of freedom and 29 denominator degrees of freedom, and it is statistically significant at the 5% significance level.
- (B) The *F* statistic has 4 numerator degrees of freedom and 29 denominator degrees of freedom, and it is not statistically significant at the 5% significance level.
- (C) The *F* statistic has 4 numerator degrees of freedom and 33 denominator degrees of freedom, and it is not statistically significant at the 5% significance level.
- (D) The *F* statistic has 8 numerator degrees of freedom and 33 denominator degrees of freedom, and it is statistically significant at the 5% significance level.
- (E) The *F* statistic has 8 numerator degrees of freedom and 33 denominator degrees of freedom, and it is not statistically significant at the 5% significance level.

20. From a population having distribution function *F,* you are given the following sample:

2.0, 3.3, 3.3, 4.0, 4.0, 4.7, 4.7, 4.7

Calculate the kernel density estimate of *F*(4), using the uniform kernel with bandwidth 1.4.

- (A) 0.31
- (B) 0.41
- (C) 0.50
- (D) 0.53
- (E) 0.63

(i) The number of claims has probability function:

$$
p(x) = {m \choose x} q^x (1-q)^{m-x}, \qquad x = 0, 1, 2, ..., m
$$

- (ii) The actual number of claims must be within 1% of the expected number of claims with probability 0.95.
- (iii) The expected number of claims for full credibility is 34,574.

Determine *q*.

- (A) 0.05
- (B) 0.10
- (C) 0.20
- (D) 0.40
- (E) 0.80
- 22. If the proposed model is appropriate, which of the following tends to zero as the sample size goes to infinity?
	- (A) Kolmogorov-Smirnov test statistic
	- (B) Anderson-Darling test statistic
	- (C) Chi-square goodness-of-fit test statistic
	- (D) Schwarz Bayesian adjustment
	- (E) None of (A) , (B) , (C) or (D)

23. The model $Y_i = \beta X_i + \varepsilon_i$ is fitted to the following observations:

X	
1.0	1.0
4.5	9.0
7.0	20.1

$$
\text{Var}\left(\varepsilon_i\right) = \sigma^2 X_i
$$

Determine the weighted least-squares estimate of β .

- (A) Less than 2.5
- (B) At least 2.5, but less than 2.8
- (C) At least 2.8, but less than 3.1
- (D) At least 3.1, but less than 3.4
- (E) At least 3.4

- (i) Losses are uniformly distributed on $(0, \theta)$ with $\theta > 150$.
- (ii) The policy limit is 150.
- (iii) A sample of payments is:

14, 33, 72, 94, 120, 135, 150, 150

Estimate θ by matching the average sample payment to the expected payment per loss.

- (A) 192
- (B) 196
- (C) 200
- (D) 204
- (E) 208

- (i) A portfolio of independent risks is divided into two classes.
- (ii) Each class contains the same number of risks.
- (iii) For each risk in Class 1, the number of claims per year follows a Poisson distribution with mean 5.
- (iv) For each risk in Class 2, the number of claims per year follows a binomial distribution with $m = 8$ and $q = 0.55$.
- (v) A randomly selected risk has three claims in Year 1, *r* claims in Year 2 and four claims in Year 3.

The Bühlmann credibility estimate for the number of claims in Year 4 for this risk is 4.6019.

Determine *r* .

- $(A) 1$
- $(B) 2$
- (C) 3
- (D) 4
- (E) 5

(i) A sample of losses is:

600 700 900

- (ii) No information is available about losses of 500 or less.
- (iii) Losses are assumed to follow an exponential distribution with mean θ .

Determine the maximum likelihood estimate of θ .

- (A) 233
- (B) 400
- (C) 500
- (D) 733
- (E) 1233

27. You are given the following model:

 $\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 (\ln X_{2t} - \ln X_{2t_0}) D_t + \beta_5 (\ln X_{3t} - \ln X_{3t_0}) D_t + \varepsilon_t$ where,

- *t* indexes the years 1979-93 and t_0 is 1990.
- *Y* is a measure of workers' compensation frequency.
- X_2 is a measure of employment level.
- X_3 is a measure of unemployment rate.
- *D_t* is 0 for $t \le t_0$ and 1 for $t > t_0$.

Fitting the model yields:
$$
\hat{\beta} = \begin{pmatrix} 4.00 \\ 0.60 \\ -0.10 \\ -0.07 \\ -0.01 \end{pmatrix}
$$

Estimate the elasticity of frequency with respect to employment level for 1992.

- (A) -0.11 (B) 0.53
- (C) 0.60
- (D) 0.90
- (E) 1.70

28. You use a Cox proportional hazards model with a data-dependent model for the baseline hazard rate to represent the mortality of two groups of two lives each. The lives in Group A die at times t_1 and t_4 , and the lives in Group B die at times t_2 and t_3 , where $t_1 < t_2 < t_3 < t_4$.

Which of the following statements is true for $t > t_1$?

- (A) The estimated cumulative hazard rate function for Group A is everywhere smaller than the estimated cumulative hazard rate function for Group B.
- (B) The estimated cumulative hazard rate function for Group A is everywhere larger than the estimated cumulative hazard rate function for Group B.
- (C) The estimated cumulative hazard rate function for Group A is identical to the estimated cumulative hazard rate function for Group B.
- (D) The estimated cumulative hazard rate function for Group A is neither everywhere smaller nor everywhere larger nor identical to the estimated cumulative hazard rate function for Group B.
- (E) The relationship between the estimated cumulative hazard rate function for Group A and the estimated cumulative hazard rate function for Group B cannot be determined without knowing the times at which the deaths occurred.

- (i) Claim counts follow a Poisson distribution with mean λ .
- (ii) Claim sizes follow a lognormal distribution with parameters μ and σ .
- (iii) Claim counts and claim sizes are independent.
- (iv) The prior distribution has joint probability density function:

$$
f(\lambda, \mu, \sigma) = 2\sigma
$$
, $0 < \lambda < 1$, $0 < \mu < 1$, $0 < \sigma < 1$

Calculate Bühlmann's *k* for aggregate losses.

- (A) Less than 2
- (B) At least 2, but less than 4
- (C) At least 4, but less than 6
- (D) At least 6, but less than 8

(E) At least 8

30. You are given the following data:

0.49 0.51 0.66 1.82 3.71 5.20 7.62 12.66 35.24

You use the method of percentile matching at the $40th$ and $80th$ percentiles to fit an Inverse Weibull distribution to these data.

Determine the estimate of θ .

- (A) Less than 1.35
- (B) At least 1.35, but less than 1.45
- (C) At least 1.45, but less than 1.55
- (D) At least 1.55, but less than 1.65
- (E) At least 1.65

31. An ARMA(*p*, *q*) model is used to represent a time series. You perform diagnostic checking to test whether the model is specified correctly.

Which of the following is false?

- (A) If the model is correctly specified, then for large displacements *k*, the residual autocorrelations \hat{r}_k are themselves uncorrelated, normally distributed random variables with mean 0 and variance 1/*T*, where *T* is the number of observations in the time series.
- (B) The Box-Pierce *Q* statistic, where $Q = T \sum_{k=1}^{n}$ *K k* $Q = T \sum \hat{r}_k^2$ 1 \hat{r}_{k}^{2} , is approximately distributed as chisquare with *K* degrees of freedom.
- (C) For low-order models, *K* equal to 15 or 20 is sufficient for the Box-Pierce *Q* statistic to be meaningful.
- (D) The Box-Pierce *Q* statistic is *approximately* chi-square distributed because the first few autocorrelations will have a variance slightly less than 1/*T* and may themselves be correlated.
- (E) If the Box-Pierce Q statistic is smaller than the $90th$ percentile of the appropriate chi-square distribution, then the model is not rejected at the 10% significance level.

- (i) The number of claims follows a Poisson distribution with mean λ .
- (ii) Observations other than 0 and 1 have been deleted from the data.
- (iii) The data contain an equal number of observations of 0 and 1.

Determine the maximum likelihood estimate of λ .

- (A) 0.50
- (B) 0.75
- (C) 1.00
- (D) 1.25
- (E) 1.50

- (i) In a portfolio of risks, each policyholder can have at most one claim per year.
- (ii) The probability of a claim for a policyholder during a year is *q* .

(iii) The prior density is
$$
\pi(q) = \frac{q^3}{0.07}
$$
, $0.6 < q < 0.8$.

A randomly selected policyholder has one claim in Year 1 and zero claims in Year 2.

For this policyholder, determine the posterior probability that $0.7 < q < 0.8$.

- (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

Insured	$X = \text{Age}$	$Y =$ Number
		of Accidents
	34	
	38	
	45	
	25	3
	21	
Total	163	

(i) The ages and number of accidents for five insureds are as follows:

(ii) $Y_1, Y_2, ..., Y_5$ are independently Poisson distributed with means $\mu_i = \beta X_i$, $i = 1,2,...,5$.

Estimate the standard deviation of $\hat{\beta}$.

- (A) Less than 0.015
- (B) At least 0.015, but less than 0.020
- (C) At least 0.020, but less than 0.025
- (D) At least 0.025, but less than 0.030
- (E) At least 0.030
- **35.** Which of the following statements regarding the ordinary least squares fit of the model $Y = \alpha + \beta X + \varepsilon$ is false?
	- (A) The lower the ratio of the standard error of the regression *s* to the mean of *Y*, the more closely the data fit the regression line.
	- (B) The precision of the slope estimator decreases as the variation of the *X*'s increases.
	- (C) The residual variance s^2 is an unbiased as well as consistent estimator of the error variance σ^2 .
	- (D) If the mean of *X* is positive, then an overestimate of α is likely to be associated with an underestimate of β .
	- (E) $\hat{\beta}$ is an unbiased estimator of β .

(i) The following is a sample of 15 losses:

11, 22, 22, 22, 36, 51, 69, 69, 69, 92, 92, 120, 161, 161, 230

- (ii) $\hat{H}_1(x)$ is the Nelson-Aalen empirical estimate of the cumulative hazard rate function.
- (iii) $\hat{H}_2(x)$ is the maximum likelihood estimate of the cumulative hazard rate function under the assumption that the sample is drawn from an exponential distribution.

Calculate $\left| \hat{H}_2(75) - \hat{H}_1(75) \right|$.

- (A) 0.00
- (B) 0.11
- (C) 0.22
- (D) 0.33
- (E) 0.44
- **37.** For a portfolio of motorcycle insurance policyholders, you are given:
	- (i) The number of claims for each policyholder has a conditional Poisson distribution.

(ii) For Year 1, the following data are observed:

Determine the credibility factor, *Z*, for Year 2.

- (A) Less than 0.30
- (B) At least 0.30, but less than 0.35
- (C) At least 0.35, but less than 0.40
- (D) At least 0.40, but less than 0.45
- (E) At least 0.45

38. You are given a random sample of observations:

0.1 0.2 0.5 0.7 1.3

You test the hypothesis that the probability density function is:

$$
f\left(x\right) = \frac{4}{\left(1+x\right)^5}, \qquad x > 0
$$

Calculate the Kolmogorov-Smirnov test statistic.

(A) Less than 0.05

- (B) At least 0.05, but less than 0.15
- (C) At least 0.15, but less than 0.25
- (D) At least 0.25, but less than 0.35
- (E) At least 0.35

39. You are given the following time-series model:

$$
y_t = 0.9 y_{t-1} + 1 + \varepsilon_t - 0.4 \varepsilon_{t-1}
$$

You are also given:

$$
y_T = 8.0
$$

$$
\hat{\varepsilon}_T = 0.5
$$

Calculate the two-period forecast, $\hat{y}_T(2)$.

- (A) 8.0
- (B) 8.2
- (C) 8.4
- (D) 8.6
- (E) 8.8

40. Which of the following statements is true?

- (A) A uniformly minimum variance unbiased estimator is an estimator such that no other estimator has a smaller variance.
- (B) An estimator is consistent whenever the variance of the estimator approaches zero as the sample size increases to infinity.
- (C) A consistent estimator is also unbiased.
- (D) For an unbiased estimator, the mean squared error is always equal to the variance.
- (E) One computational advantage of using mean squared error is that it is not a function of the true value of the parameter.

****END OF EXAMINATION****

Course 4, Fall 2004

FINAL ANSWER KEY

