MATH461, Winter 2003. **Solutions**

1. [**The problem was not seen. Evaluation of "difficult" probabilities like γ were discussed in class. The law of total probability and Bayes' rule were studied in depth.**]

   (a) The next two services after deuce can result in Rod's success with probability p^2, in Fred's success with probability $(1-p)^2$ and in the new deuce with probability $2p(1-p)$. (Either Rod gains the advantage and loses it, or Fred gains the advantage and loses it.) Hence, by the law of total probability,

   $$\gamma = p^2 \times 1 + (1-p)^2 \times 0 + 2p(1-p) \times \gamma$$

   and

   $$\gamma = \frac{p^2}{1-2p(1-p)}.$$

   (b) Let T and N be events "Fred trained the day before" and "Fred did NOT train". Then

   $$\gamma(T) = P(\text{Rod wins}|T) = \frac{0.49^2}{1-2\cdot 0.49 \cdot 0.51} = 0.48;$$

   $$\gamma(N) = P(\text{Rod wins}|N) = \frac{0.51^2}{1-2\cdot 0.51 \cdot 0.49} = 0.52.$$

   Hence, by the law of total probability

   $$P(\text{Rod wins}) = P(\text{Rod wins}|T)P(T) + P(\text{Rod wins}|N)P(N)$$

   $$= 0.48 \cdot 0.75 + 0.52 \cdot 0.25 = 0.36 + 0.13 = 0.49.$$

   (c) Using Bayes' rule we have

   $$P(T|\text{Rod wins}) = \frac{P(\text{Rod wins}|T)P(T)}{P(\text{Rod wins})} = \frac{0.48 \cdot 0.75}{0.49} = 0.73.$$

   We see that $0.73 < 0.75$. This is natural since the chances that Fred had trained the day before decrease after he loses the game.

2. [**Standard, but not seen. Other similar problems describing transmission of signals were discussed in tutorials.**]

   (i) If $s = 0$ then the signal needs to be transmitted more than once if $Z \in [0.4, \ 0.6]$. Let X be the number of transmissions till the message can be decoded without errors. Now

   $$P(X > 1|s = 0) = P(0.4 \le Z \le 0.6) = \frac{0.2}{1.2} = 1/6.$$

   Similarly

   $$P(X > 1|s = 1) = P(-0.6 \le Z \le -0.4) = \frac{0.2}{1.2} = 1/6$$

and these two probabilities coincide.

(ii) Using the hint we obtain

$$E[X] = \sum_{k=1}^{\infty} kP(X = k) = \frac{5/6}{(5/6)^2} = \frac{6}{5}.$$

Let Y be the number of messages from B to A asking for retransmissions (per one correctly decoded signal). Then $Y = X - 1$ and

$$E[Y] = E[X] - 1 = 6/5 - 1 = 1/5.$$

Thus the total expected number of messages per one correctly decoded signal is

$$E[X] + E[Y] = 7/5.$$

(iii) Let X be the number of coupled transmissions (per one correctly decoded signal). Since $X > 1$ means that neither of two messages can be decoded correctly,

$$P(X > 1) = (1/6)^2 = \frac{1}{36};$$

$$P(X = k) = (1/36)^{k-1}(35/36), \quad k = 1, 2, \ldots$$

and

$$E[X] = \frac{35/36}{(35/36)^2} = \frac{36}{35}.$$

If Y is, as previously, the number of messages from B to A asking for retransmission then $Y = X - 1$ and $E[Y] = \frac{1}{35}$.

Therefore, the expected number of messages sent by either computer per one correctly decoded signal is

$$2 \times \frac{36}{35} + \frac{1}{35} = \frac{73}{35}.$$

3. **[This problem was not seen, but several similar questions concerning quality of products were set for independent work at home.]**

   (i) Standard estimators of the mean and of the varience are given by formulae

$$\hat{\mu} = \bar{X} = \frac{1}{10}\sum_{i=1}^{10} x_i = 29.9;$$

$$\hat{\sigma}^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{X})^2 = 16.32.$$

   If $p = const$ then $X \sim Bin(100, p)$ and the estimator for p is given by

$$\hat{p} = \frac{\bar{X}}{100} \approx 0.3.$$

Since $n = 100$ is large and $\hat{p} = 0.3$ is moderate we can use the normal approximation to the binomial distribution:

$$X \sim N(n\hat{p}, n\hat{p}(1-\hat{p})) = N(30, 21).$$

Therefore, the estimator of $Var(X)$ is 21 which is not too far from $\hat{\sigma}^2 = 16.32$ obtained earlier. Hence all our assumptions make sense.

(ii) Now

$$P(X \le 20) = P(Z \le \frac{20-30}{\sqrt{21}}) = P(Z \le -2.18) \approx 0.015,$$

$$P(X \ge 37) = P(Z \ge \frac{37-30}{\sqrt{21}}) = P(Z \ge 1.53) \approx 0.063,$$

where $Z \sim N(0,1)$ is the standard normal RV.

(iii) Using observations y_i we get $\hat{\mu} = 29.9$, $\hat{\sigma}^2 = 32.28$, $\hat{p} \approx 0.3$. If we used the normal approximation to RV

$$Y = \text{"number of dead seeds in a packet of 100"}$$

we would conclude that $Y \sim N(30, 21)$, and $Var(Y) = 21$ as previously and the probabilities evaluated in (ii) do not change. At the same time 21 is far away from $\hat{\sigma}^2 = 32.28$. This is a signal that our assumptions probably don't hold. Moreover, we have observations $y_2 = 20$ and $y_6 = y_7 = 37$ which have small probabilities, if all previous hypotheses hold. Therefore, having data $\{y_i\}$ it is maybe worth rejecting hypothesis that p is constant for each packet: if the seeds in different packets come from different years then hypothesis $p = const$ is unlikely to hold.

4. [**Not seen; maximum likelihood estimators were discussed in class for other models as well as plotting and hypotheses testing.**]

(a) If CDF is $F(x) = 1 - e^{-\lambda x}$ then the density function is

$$f(x) = \frac{dF(x)}{dx} = \lambda e^{-\lambda x}, \quad x \ge 0.$$

Hence, the likelihood function is

$$L_x(\lambda) = f(x_1)f(x_2)\dots f(x_n) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}.$$

The maximum $\max_{\lambda \ge 0} L_x(\lambda)$ is attained at the same point as the maximum of logarithm $\max_{\lambda \ge 0} \ln\ L_x(\lambda)$ which can be found by differentiating:

$$\frac{d}{d\lambda} \ln\ L_x(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0.$$

Hence,

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{n}{\bar{X}}$$

is the maximum likelihood estimator for λ.

(b) $\bar{X} = 0 \times 12 + 1 \times 8 + 2 \times 7 + 4 \times 8 = 54$ and

$$\hat{\lambda} = \frac{35}{54} = 0.648.$$

(c) The sample cumulative distribution is given by the formula

$$\hat{F}(x_i) = \frac{\text{No. of observations} \leq x_i}{n+1}.$$

In our case, $n = 35$; $x_i = 0, 1, 2, 4$.

Since

$$-\ln(1 - F(x)) = \lambda x$$

it is worth plotting $-\ln(1 - \hat{F}(x_i))$ against x_i. The results of calculations are presented in the table:

x_i	0	1	2	4
$\hat{F}(x_i)$	0.333	0.556	0.750	0.97
$-\ln(1 - \hat{F}(x_i))$	0.40	0.81	1.39	3.51

and the graph looks like following

We can see that the points are approximately on a straight line meaning that the data fit exponential model.

The tangent of inclination of the straight line gives the value of $\lambda \approx 0.8$ which is not far from $\hat{\lambda} = 0.648$.

5. [**Pareto distribution was not seen; CDF, densities and expectations were discussed in depth; calculations of series and parallel systems were set for independent homework.**]

(a) We have, since $f(x) = 0$, $x < 0$:

$$F(x) = \int_{-\infty}^{x} f(x)dx = 0 \text{ if } x < 0.$$

For $x \geq 0$,

$$F(x) = \int_0^x \left(\frac{\delta}{\theta}\right)\left(1+\frac{x}{\theta}\right)^{-\delta-1} dx.$$

Put $y = 1 + \frac{x}{\theta}$, $\frac{dy}{dx} = \frac{1}{\theta}$:

$$F(x) = \int_1^{1+x/\theta} \frac{\delta}{\theta}\frac{1}{y^{\delta+1}}\theta dy = \delta\left[\frac{-y^{-\delta}}{\delta}\right]_1^{1+\frac{x}{\theta}} = 1 - \left(1+\frac{x}{\theta}\right)^{-\delta}.$$

Also,

$$E[X] = \int_0^\infty x\left(\frac{\delta}{\theta}\right)\left(1+\frac{x}{\theta}\right)^{-\delta-1} dx = \int_1^\infty \theta(y-1)\left(\frac{\delta}{\theta}\right)\frac{1}{y^{\delta+1}}\theta dy$$

$$= \delta\theta\left\{\int_1^\infty \frac{1}{y^\delta}dy - \int_1^\infty \frac{1}{y^{\delta+1}}dy\right\} = \delta\theta\left\{\left[\frac{y^{-\delta+1}}{-\delta+1}\right]_1^\infty - \left[\frac{y^{-\delta}}{-\delta}\right]_1^\infty\right\}$$

$$= \delta\theta\left\{\frac{1}{\delta-1} - \frac{1}{\delta}\right\} = \frac{\delta\theta(\delta-\delta+1)}{(\delta-1)\delta} = \frac{\theta}{\delta-1}.$$

(b) We have

$$P(Y > y) = P(X_1 > y \cap X_2 > y) = [1-F(y)]^2 = \left(1+\frac{y}{\theta}\right)^{-2\delta}.$$

Hence

$$G(y) = P(Y \leq y) = 1 - \left(1+\frac{y}{\theta}\right)^{-2\delta}$$

and

$$g(y) = \frac{dG(y)}{dy} = \left(\frac{2\delta}{\theta}\right)\left(1+\frac{y}{\theta}\right)^{-2\delta-1}.$$

Clearly Y is Pareto with parameters 2δ and θ. Thus $E[Y] = \frac{\theta}{2\delta-1}$.
According to the definition, hazard function is

$$h(y) = \frac{g(y)}{1-G(y)} = \frac{\left(\frac{2\delta}{\theta}\right)\left(1+\frac{y}{\theta}\right)^{-2\delta-1}}{\left(1+\frac{y}{\theta}\right)^{-2\delta}} = \frac{\left(\frac{2\delta}{\theta}\right)}{1+\frac{y}{\theta}}.$$

Since $h(y)$ decreases, this piece is less likely to fail in the next increment of time $(y, y+\Delta y)$ than it would be in an interval of the same length at an earlier age. To put it defferently, the piece is improving with age. Such models describe "young" systems.

(c) In this case

$$P(W \leq w) = P(w) = P(X_3 \leq w \cap X_4 \leq w) = [F(w)]^2$$

$$= \left[1 - \left(1 + \frac{w}{\theta}\right)^{-\delta}\right]^2 = 1 - 2\left(1 + \frac{w}{\theta}\right)^{-\delta} + \left(1 + \frac{w}{\theta}\right)^{-2\delta}, \qquad w > 0.$$

Hence

$$p(w) = \frac{dP(w)}{dw} = \frac{2\delta}{\theta}\left(1 + \frac{w}{\theta}\right)^{-\delta-1} - \frac{2\delta}{\theta}\left(1 + \frac{w}{\theta}\right)^{-2\delta-1}$$

$$= (2\delta/\theta)\left(1 + \frac{w}{\theta}\right)^{-\delta-1}\left[1 - \left(1 + \frac{w}{\theta}\right)^{-\delta}\right].$$

We have

$$E[W] = \frac{2\delta}{\theta}\int_0^\infty w\left(1 + \frac{w}{\theta}\right)^{-\delta-1} dw - \frac{2\delta}{\theta}\int_0^\infty w\left(1 + \frac{w}{\theta}\right)^{-2\delta-1} dw$$

$$= 2E[X] - E[Y] = \frac{2\theta}{\delta - 1} - \frac{\theta}{2\delta - 1}.$$

(d) Units in parallel are also referred as redundant units. Redundancy is a very important aspect of system design and reliability in that adding redundancy is one of several methods of improving system reliability. In a parallel configuration, the component with the highest reliability has the biggest effect on the system's reliability, since the most reliable component is the one that will most likely fail last. Lastly, as the number of parallel components increases, the system's reliability increases.

6. **[Random generators were discussed in class as well the Law of Large Numbers. Properties of the Cauchy RV were not discussed.]**

(i) The period of a random generator is the smallest integer p such that $x_p = x_0$. Good generators have large values of the period.

(ii) If $X \sim U(0,1)$ then $Y = F^{-1}(X)$ is the new RV having the given CDF $F(y)$.

(iii) Firstly, CDF of Cauchy RV is given by

$$F(y) = \int_{-\infty}^{y} \frac{1}{\pi(1+y^2)} dy = \frac{1}{\pi}[\tan^{-1}(y) + \frac{\pi}{2}] = \frac{1}{2} + \frac{\tan^{-1}(y)}{\pi}.$$

Now, if $X_i \sim U(0,1)$ then $Y_i = \tan[(X_i - 0.5)\pi]$ is the sequence of iid Cauchy RVs.

For x_i : 0.2311; 0.6068 we have y_i : $-$ 1.126; 0.349.

(iv) The average values are presented below.

$\bar{y}_n$	-1.126	-0.3885	-0.274	0.498	0.614	0.489	-2.036
$\bar{y}_n$	-1.583	-1.426	-1.246	-1.014	-0.597	-0.479	-0.561
$\bar{y}_n$	-0.544	-0.206	0.027	0.009	0.160	-0.119	

There is no evidence that $\bar{y}_i$ stabilizes. Indeed, among the last ten observations we have the values -1.014 and +0.160 corresponding to large oscillations. This is natural since Cauchy RV has no mathematical expectation, and according to the Law of Large Numbers, the sequence $\bar{y}_n$ converges to the expectation, if it exists.

(v) CDF of RV Z is given by

$$F(z) = \int_0^z e^{-z} dz = 1 - e^{-z}.$$

Now, if $X_i \sim U(0,1)$ then $Z_i = -\ln(1 - X_i)$ is the sequence of exponential RV required.

For x_i : 0.2311; 0.6068 we have z_i : 0.263; 0.933. The average values are presented below.

0.263	0.598	0.621	1.020	1.103	1.021	0.878
0.984	0.939	0.941	0.998	1.127	1.144	1.076
1.039	1.145	1.223	1.185	1.240	1.181	

In this case, $\bar{z}_i$ seems to converge to the limiting value around 1.0. Again, this is natural since $E[Z] = 1$ and, according to the Law of Large Numbers, the sequence $\bar{z}_n$ converges, in probability, to $E[Z]$.

7. [**The problem was not seen. General Goel-Okumoto model was presented on the lecture; all the questions are standard.**]

(i) Firstly we must determine $\mu(t)$:

$$\mu(t) = \mu(0) + \int_0^t \lambda(s) ds = 0 + \int_0^t a s^{b-1} ds = \frac{a}{b} t^b.$$

Hence $\mu(t) = 2t^{1/2} = 2\sqrt{t}$.

Now the probability of two or more failures during the interval $[1, T)$ equals $1 - p$, where

$$p = P(X(1,T) = 0) + P(X(1,T) = 1) = e^{-(2\sqrt{T}-2)}[1 + 2\sqrt{T} - 2]$$

and we must determine the minimal $T > 1$ such that the last expression is less than 0.5.

Try $T = 2$: $p = 0.8538$.

Try $T = 3$: $p = 0.5699$.

Try $T = 4$: $p = 0.4060$.
Therefore, $T = 4$ is the answer.

(ii)

$$E[X(0,1)] = \mu(1) - \mu(0) = 2;$$
$$E[X(1,2)] = \mu(2) - \mu(1) = 2\sqrt{2} - 2 \approx 0.828.$$

Therefore, $E[X(0,1)] > 2E[X(1,2)]$ as required. Usually, during the debugging period, in the later time intervals, less number of failures are found. Thus the model considered is plausible.

Let A and B $\in \{0,1,2\}$ be the numbers of failures met by the first and second specialist correspondingly. Then

$$P(A=0) = \exp[-\mu(1)+\mu(0)] = e^{-2} = 0.135;$$
$$P(A=1) = e^{-2} \times 2 = 0.271;$$
$$P(A=2) = 1 - e^{-2}[1+2] = 1 - 0.406 = 0.594,$$

since the last probability coincides with the probability to have more than 1 failure in the first hour.

$$P(B=0) = P(A<2) \times [-\mu(2)+\mu(1)] + P(A=2)$$
$$= 0.406e^{-0.828} + 0.594 = 0.771,$$

since if $A=2$ then definitely $B=0$.

$$P(B=1) = P(A=0)\exp[-\mu(2)+\mu(1)][\mu(2)-\mu(1)]$$
$$+P(A=1)\{1-\exp[-\mu(2)+\mu(1)] \times [\mu(2)-\mu(1)]\}$$
$$= 0.135e^{-0.828}0.828 + 0.271\{1-e^{-0.828}0.828\} = 0.049 + 0.173 = 0.222,$$

since in the second case, when $A=1$, one or more failures in the second hour imply that $B=1$.

$$P(B=2) = P(A=0)\{1-e^{-0.828}(1+0.828)\} = 0.135 \times 0.201 = 0.027.$$

The expected profit of each specialist can be calculated in the usual way:
A: $30P(A=1) + 60P(A=2) = 43.76$;
B: $30P(B=1) + 60P(B=2) = 8.28$.

(iii) It is well known for Poisson RV that

$$E[X(0,T)] = \mu(T) - \mu(0).$$

If $\lambda(t) = ce^{-dt}$ then

$$E_2 = \mu(T) = 0 + \int_0^T ce^{-ds}ds = \frac{c}{d}[1-e^{-dT}].$$

Similarly, $E_1 = \frac{a}{b}T^b$. Therefore, $\lim_{T\to\infty} E_1 = +\infty$ and $\lim_{T\to\infty} E_2 = \frac{c}{d}$. Thus, the first model predicts unbounded increase of failures (on average) for increasing time horizon, while the second model predicts that the mean total number of failures is bounded. The first model with $\lambda(t) = at^{b-1}$ is not adequate for large intervals $[0,T]$ if the specialists working with it correct arising errors after each failure. In this situation, second model with $\lambda(t) = ce^{-dt}$ seems to be more satisfactory. At the same time, for short time intervals $[0,T]$, both models make sense and can be adequate for different types of software. Lastly, if mistakes in programs are not corrected then the first model is definitely more accurate.