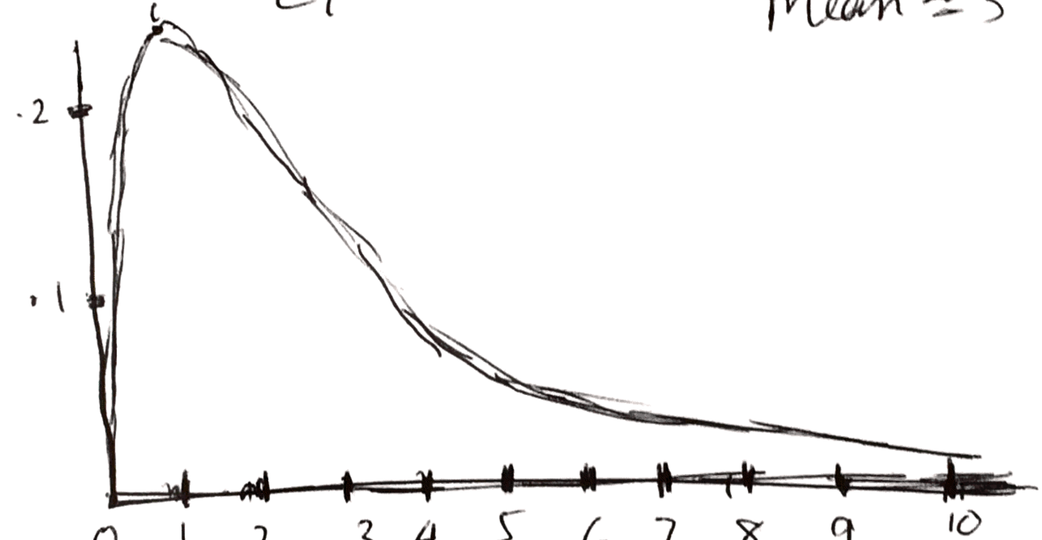
Sol. Probset 3

1. We have 4 probabilities known. So the degree of freedom is 3.

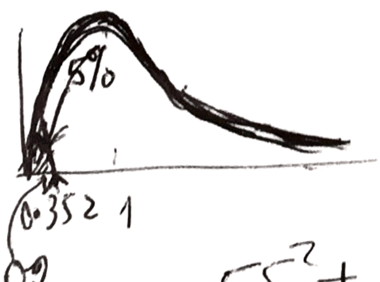
$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$ is approximately χ^2 d.f. 3.

Mean = 3 mode = 1



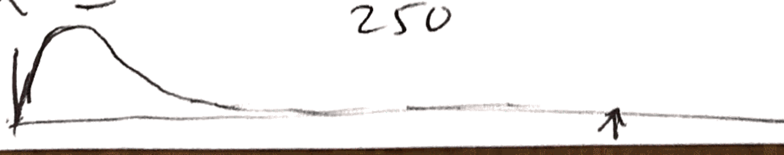
a) $$\chi^2 = \frac{1^2 + 2^2 + 6^2 + 7^2}{250} = \frac{50}{250} = 0.2$$

This looks too low. The lower 5% threshold is 0.352. The observed χ^2 is lower than that.



b) $$\chi^2 = \frac{55^2 + 31^2 + 70^2 + 46^2}{250} = 44.008$$

Too high! P value of the order of 10^{-9}.



$$\frac{e^{-\chi^2/2}(\chi^2)^{+1/2}}{2^{3/2}\Gamma(\frac{3}{2})} = \frac{e^{-\chi^2/2}(\chi^2)^{1/2}}{2^{3/2}\frac{1}{2}\sqrt{\pi}} = \frac{e^{-\chi^2/2}(\chi^2)^{1/2}}{\sqrt{2\pi}}$$

$$\int_{\chi_\nu^2}^{\infty} d(\chi^2)\,\frac{e^{-\chi^2/2}(\chi^2)^{1/2}}{\sqrt{2\pi}} = 2\int_{\frac{\chi_\nu^2}{2}}^{\infty} dy\,\frac{e^{-y}(2y)^{1/2}}{\sqrt{2\pi}} \qquad \chi^2/2 = y$$

$$= \frac{2}{\sqrt{\pi}}\int_{\frac{\chi_\nu^2}{2}}^{\infty} e^{-y}y^{1/2}\,dy \approx \frac{2}{\sqrt{\pi}}\,e^{-\frac{\chi_\nu^2}{2}}\sqrt{\chi_\nu^2} = \frac{2}{\sqrt{\pi}}\,e^{-22}\sqrt{44} \sim 10^{-9}$$

[2×10^{-9} from the approx calc., 1.5×10^{-9} according to Matlab]

Because $$\int_{y_0}^{\infty} e^{-y}y^{1/2}\,dy = -\int_{y\in[y_0,\infty]} d(e^{-y})\,y^{1/2} = -\int_{y\in[y_0,\infty]} d(e^{-y}y^{1/2}) + \int_{y\in[y_0,\infty]} e^{-y}\,d(y^{1/2})$$

$$= +e^{-y_0}y_0^{1/2} - \frac{1}{2}\int_{y_0}^{\infty}\frac{e^{-y}}{\sqrt{y}}\,dy$$ ← ignore for large y_0

c) If we scale things up by a factor of λ, $(O_i - E_i)^2 \sim \lambda^2$ while $E_i \sim \lambda$. So χ^2 scales like λ. We expect the χ^2 in 40 colony expt to be order 1 (on the average 3). Multiplying by 25 scales it upto a number order 10, making it an unlikely value.

2. a) I got the MLE (Least Sq) estimates to be $\hat{w}_0 = -3.2584$

$\hat{w}_1 = 0.0427$

$\hat{\sigma}^2 = 1.7 \times 10^{-2}$

b) I will pretend $P(w_0) = \sqrt{\frac{\epsilon}{\pi}} e^{-\epsilon w_0^2}$

When $\epsilon \to 0$ this becomes a very wide distribution of width $\frac{1}{\sqrt{\epsilon}}$ with nearly constant ~~value~~ probability density for $|w_0| \ll \frac{1}{\sqrt{\epsilon}}$.

You can choose a box of length L, $[-\frac{L}{2}, \frac{L}{2}]$ with pdf $\frac{1}{L}$. Many different choices would give the same answer in the "wide" prior limit, one data starts to constrain w_0.

$$P(\underline{w}) = \sqrt{\frac{\epsilon}{\pi}} e^{-\epsilon w_0^2} \sqrt{\frac{1}{2\pi}} e^{-\frac{w_1^2}{2}}$$

(w_0, w_1)

$P(\underline{w}) = \mathcal{N}(\underline{w} \mid \underline{w}_0, \underline{V}_0)$

$\underline{w}_0 = (0, 0)$ $\quad V_0 = \begin{pmatrix} \frac{1}{2\epsilon} & 0 \\ 0 & 1 \end{pmatrix}$

c) Let us take the posterior:

$$P(w_0, w_1 | D, \hat\sigma^2) = e^{-\sum_i (y_i - w_0 - w_1 x_i)^2 / 2\hat\sigma^2} \, e^{-\epsilon w_0^2} \, e^{-\frac{w_1^2}{2}}$$

Let $x_i' = x_i - \bar{x}$
$y_i' = y_i - \bar{y}$

$$\sum_i (y_i - w_1 x_i - w_0)^2$$
$$= \sum_i (\bar{y} - w_1 \bar{x} - w_0 + y_i' - w_1 x_i')^2$$
$$= N(\bar{y} - w_1 \bar{x} - w_0)^2 + \sum_i (y_i' - w_1 x_i')^2$$
$$+ (\bar{y} - w_0 - w_1 \bar{x}) \underbrace{\sum_i (y_i' - w_1 x_i')}_{0}$$

So $P(w_0, w_1 | D, \hat\sigma^2) \propto e^{-\epsilon w_0^2 - (w_0 - \bar{y} + w_1 \bar{x})^2 / 2\hat\sigma^2}$
$$e^{-\frac{\sum_i (y_i' - w_1 x_i')^2}{2\hat\sigma^2} - \frac{w_1^2}{2}}$$

When $\epsilon \to 0$, the integral over w_0 becomes essentially $\int dw_0 \, e^{-(w_0 - \bar{y} + w_1 \bar{x})^2 / 2\hat\sigma^2}$ and is independent of w_1. So the marginalized posterior of w_1 is

$$P(w_1 | D, \hat\sigma^2) \propto e^{-\frac{\sum (y_i' - w_1 x_i')^2}{2\hat\sigma^2} - \frac{w_1^2}{2}}$$

$$\propto e^{\frac{\sum x_i' y_i'}{\hat{\sigma}^2} w_1 - \frac{1}{2}\left(\frac{\sum_i x_i'^2}{\hat{\sigma}^2} + 1\right) w_1^2}$$

If we complete the square and ignore the w_1 independent term

$$P(w_1 | \mathcal{D}, \hat{\sigma}^2) \propto e^{-\frac{1}{2}\left(\frac{N S_{xx}}{\hat{\sigma}^2} + 1\right)\left(w_1 - \frac{N S_{xy}}{N S_{xx} + \hat{\sigma}^2}\right)^2}$$

$$S_{xx} = \sum_i (x_i - \bar{x})^2 / N = \sum_i x_i'^2 / N$$ → ~~empirical~~ sample variance of x

$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) / N = \frac{\sum_i x_i' y_i'}{N}$$ → sample covariance

So $$E(w_1 | \mathcal{D}, \hat{\sigma}^2) = \frac{S_{xy}}{S_{xx} + \hat{\sigma}^2 / N} = 4.27 \times 10^{-2}$$

(the $\hat{\sigma}^2/N$ in the denominator causes the shrinkage effect due to the prior)

$$\mathrm{Var}(w_1 | \mathcal{D}, \hat{\sigma}^2) = \frac{\hat{\sigma}^2}{N S_{xx} + \hat{\sigma}^2} = 1.1423 \times 10^{-5}$$

~~Note that the posterior mean is about the same as the LS estimate but order magnitude smaller~~

Note that the mean $(\approx 4\times10^{-2})$ is hardly affected by the prior, and the std $(\approx 3\times10^{-3})$ is an order of magnitude smaller.

d) $P\{\text{mean} - 1.96\,\text{std} < w_1 < \text{mean} + 1.96\,\text{std}\} = 95\%$.

Since the posterior is just a Gaussian.
That interval is $[3.6\times10^{-2} \;\; 4.93\times10^{-2}]$.