CW1/1. **Conditional probabilities.**

During a game of poker you want to compute the probability that your opponent has a pair of aces. Consider the simplified problem of a minimal pack of cards, containing just 4 cards of which exactly 2 are aces (spades and clubs). Your opponent’s hand consists of just 2 cards (and you have not seen your hand yet).

(a) What is the probability that your opponent has a pair of aces? Consider counting the number of possible hands in the game.

(b) What is the probability that your opponent has a pair of aces, knowing that she has an ace (whose suit is however unknown to you)? Use a counting argument to show that the space of possible hands is now reduced.

(c) What is the probability that your opponent has a pair of aces, knowing that she has the ace of spades? How is this different from the probability above? Explain why obtaining extra information about the suit of the ace changes the probability of a pair.

CW1/2. **A problem in medical evidence.**

A batch of chemistry undergraduates are screened for a dangerous medical condition called *Bacillus* *Bacillus* *Bacillus* *Bacillus* (BB). The incidence of the condition in the population (i.e., the probability that a randomly selected person has the disease) is estimated at about 1%. If the person has BB, the test returns positive 95% of the time. There is also a known 5% rate of false positives, i.e. the test returning positive even if the person is free from BB. One of your friends takes the test and it comes back positive. Here we examine whether your friend should be worried about her health.

(a) Translate the information above in suitably defined conditional probabilities. The two relevant propositions here are whether the test returns positive (denote this with a + symbol) and whether the person is actually sick (denote this with the symbol BB = 1. Denote the case when the person is healthy as BB = 0).

(b) Compute the conditional probability that your friend is sick, knowing that she has tested positive, i.e. find \( P(BB = 1|+) \).

(c) Imagine screening the general population for a very rare disease, whose incidence in the population is \(10^{-6}\) (i.e., one person in a million has the disease on average, i.e. \( P(BB = 1) = 10^{-6}\)). What should the reliability of the test (i.e., \( P(+|BB = 1) \)) be if we want to make sure that the probability of actually having the disease after testing positive is at least 99%? Assume first that the false positive rate \( P(+|BB = 0) \) (i.e., the probability of testing positive while healthy), is 5% as in part (a). What can you conclude about the feasibility of such a test?

(d) Now we write the false positive rate as \( P(+|BB = 0) = 1 - P(-|BB = 0) \). It is reasonable to assume (although this is not true in general) that \( P(-|BB = 0) = P(+|BB = 1) \), i.e. the probability of getting a positive result if you have the disease is the same as the probability of getting a negative result if you don’t have it. Find the requested reliability of the test (i.e., \( P(+|BB = 1) \)) so that the probability of actually having the disease after testing positive is at least 99% in this case. Comment on whether you think a test with this reliability is practically feasible.

CW1/3. **Statistics in the kitchen.**

A pan contains 10 ravioli, of which 9 are filled with mushrooms and one with ricotta. You put in the pan a further raviolo filled with mushrooms and cover with an opaque lid. Then you randomly draw a raviolo, eat it and discover that it is filled with mushrooms. After this procedure, the pan is again in the same state as before. What is now the probability that the next raviolo drawn will be filled with mushrooms?

On a different night, you cook a pan of mixed mushrooms and ricotta ravioli (in equal proportions). One last raviolo remains in your plate, which could be either mushrooms or ricotta. Your friend tosses into your plate her last raviolo, which she tells you is a mushrooms-filled one. Then you mix
the two ravioli randomly, pick one and realize it's mushrooms. What is now the probability that the last raviolo in your plate is mushrooms?
CW2/1. **Coin tossing.** This problem extends Example 20 in the handout.

You flip a coin $n = 10$ times and you obtain 8 heads.

(a) What is the likelihood function for this measurement? Identify explicitly what are the data and what is the free parameter you are trying to estimate.

(b) What is the Maximum Likelihood Estimate for the probability of obtaining heads in one flip, $p$?

(c) Approximate the likelihood function as a Gaussian around its peak and derive the $1\sigma$ confidence interval for $p$. How would you report your result for $p$?

(d) With how many $\sigma$ confidence can you exclude the hypothesis that the coin is fair? (Hint: compute the distance between the MLE for $p$ and $p = 1/2$ and express the result in number of $\sigma$).

(e) You now flip the coin 1000 times and obtain 800 heads. What is the MLE for $p$ now and what is the $1\sigma$ confidence interval for $p$? With how many $\sigma$ confidence can you exclude the hypothesis that the coin is fair now?

---

CW2/2. **Counting experiment.**

(a) An experiment counting particles emitted by a radioactive decay measures $r$ particles per unit time interval. The counts are Poisson distributed. If $\lambda$ is the average number of counts per per unit time interval, write down the appropriate probability distribution function for $r$.

(b) Now we seek to determine $\lambda$ by repeatedly measuring for $M$ times the number of counts per unit time interval. This series of measurements yields a sequence of counts $\tilde{r} = \{\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, ..., \tilde{r}_M\}$. Each measurement is assumed to be independent. Derive the combined likelihood function for $\lambda$, $\mathcal{L}(\lambda) = P(\tilde{r}|\lambda)$, given the measured sequence of counts $\tilde{r}$.

(c) Use the Maximum Likelihood Principle applied to the the log likelihood $\ln \mathcal{L}(\lambda)$ to show that the Maximum Likelihood estimator for the average rate $\lambda$ is just the average of the measured counts, $\bar{r}$, i.e.

$$
\lambda_{\text{ML}} = \frac{1}{M} \sum_{i=1}^{M} \tilde{r}_i.
$$

(d) By considering the Taylor expansion of $\ln \mathcal{L}(\lambda)$ to second order around $\lambda_{\text{ML}}$, derive the Gaussian approximation for the likelihood $\mathcal{L}(\lambda)$ around the Maximum Likelihood point (see Eq. (63) in the handout), and show that it can be written as

$$
\mathcal{L}(\lambda) \approx L_0 \exp \left(-\frac{1}{2} \frac{M}{\lambda_{\text{ML}}} (\lambda - \lambda_{\text{ML}})^2 \right),
$$

where $L_0$ is a normalization constant.

(e) Compare with the equivalent expression for $M$ Gaussian-distributed measurements to show that the variance $\sigma^2$ of the Poisson distribution is given by $\sigma^2 = \lambda$.

---

CW2/3. **Gaussian measurements with different variance.** This problem generalizes Examples 22 and 27 in the handout to the case where the measurements have different uncertainties among them.

You measure the flux $F$ of photons from a laser source using 4 different instruments and you obtain the following results (units of $10^4$ photons/cm$^2$):

$$
34.7 \pm 5.0, \quad 28.9 \pm 2.0, \quad 27.1 \pm 3.0, \quad 30.6 \pm 4.0.
$$

(1)

(a) Write down the likelihood for each measurement, and explain why a Gaussian approximation is justified in this case.

(b) Write down the total likelihood for the combination of the 4 measurements.
(c) Find the MLE of the photon flux, $F_{\text{ML}}$, and show that it is given by:

$$F_{\text{ML}} = \sum_i \frac{\hat{n}_i}{\hat{\sigma}_i^2 / \sigma^2},$$

where

$$\frac{1}{\sigma^2} = \sum_i \frac{1}{\hat{\sigma}_i^2}.$$  

(d) Compute $F_{\text{ML}}$ from the data above and compare it with the sample mean.

(e) Find the 1σ confidence interval for your MLE for the mean, and show that it is given by:

$$\left( \sum_i \frac{1}{\hat{\sigma}_i^2} \right)^{-1/2}$$

Evaluate the confidence interval for the above data. How would you summarize your measurement of the flux $F$?
CW1.1. Conditional probabilities.

(a) Given our minimal pack of cards, there are \( \binom{4}{2} = 6 \) possible pairs that can be formed. Of those pairs, 1 is a pair of aces, hence in case (a) the probability is 1/6.

(b) If however we know that our opponent has at least one ace (whose suit is unknown), the number of possible pairs is reduced to 5 (the 5 pairs containing at least one ace of any suit), hence in case (b) the probability is 1/5.

The same result can be found by using conditional probabilities. Let \( a \) denote the statement “your opponent has one ace” and \( aa \) the statement “your opponent has two aces”. Then

\[
P(aa|a) = \frac{P(aa, a)}{P(a)} = \frac{1/6}{5/6} = \frac{1}{5},
\]

where we have used \( P(aa, a) = P(aa) = 1/6 \), since the probability of having two aces and one ace is the same as having just two aces.

(c) In this case, knowledge of the suit of the ace further reduces the space of possible pairs to 3. Hence the probability is now 1/3. Knowledge of the suit does reduce the number of possible pairs that we should consider in evaluating the probability, hence this is relevant information. The argument that it should not matter because after all an ace has to be of a particular suit is flawed: once the suit is specified, the conditional probability changes. This also follows from consideration of the conditional probability. Let \( as \) denote the proposition “your opponent has the ace of spades”. Then

\[
P(aa|as) = \frac{P(aa, as)}{P(as)} = \frac{1/6}{3/6} = \frac{1}{3},
\]

where it is clear that the result is different from (b) because in the denominator above \( P(as) \neq P(a) \).

CW1.2. A problem in medical evidence.

Let \( BB = 1 \) denote the proposition that your friend has the virus, and \( BB = 0 \) that she does not. We use + (−) to denote the test returning a positive (negative) result.

(a) We know from the reliability of the test that

\[
P(+|BB = 1) = 0.95 \quad (3)
\]

\[
P(+|BB = 0) = 0.05 \quad \text{hence} \quad (4)
\]

\[
P(−|BB = 0) = 0.95. \quad (5)
\]

Given that 1% of the population has the virus, the probability of being one of them (before taking the test) is \( P(BB = 1) = 0.01 \), while \( P(BB = 0) = 0.99 \).

(b) The probability of your friend having the virus after she has tested positive is thus

\[
P(BB = 1|+) = \frac{P(+|BB = 1)P(BB = 1)}{P(+)} \quad (6)
\]

We can compute the denominator as follows:

\[
P(+) = P(+|BB = 1)P(BB = 1) + P(+|BB = 0)P(BB = 0) = 0.95 \cdot 0.01 + 0.05 \cdot 0.99 = 0.059. \quad (7)
\]

Therefore the probability that your friend has the virus is much less than 95%, namely

\[
P(BB = 1|+) = \frac{0.95 \cdot 0.01}{0.059} = 0.16 = 16%. \quad (8)
\]
(c) From the above, we have that
\[
P(BB = 1|+) = \frac{P(+|BB = 1)P(BB = 1)}{P(+|BB = 1)P(BB = 1) + P(+|BB = 0)P(BB = 0)}.
\] (9)

We want to achieve 99% probability that the person has BB given that they tested positive, i.e., \(P(BB = 1|+) = 0.99\), and we need to solve the above equation for the reliability, i.e. \(P(+|BB = 1)\).

We first assume that \(P(+|BB = 0) = 0.05\), as in part (a). Since \(P(BB = 1) = 10^{-6}\), it follows that \(P(BB = 0) = 1 - P(BB = 1) \approx 1.\) Then Eq. (9) becomes

\[
0.99 = \frac{P(+|BB = 1)10^{-6}}{P(+|BB = 1)10^{-6} + 0.05 \times 1} \approx 10^{-6}/0.05.
\] (10)

It is clear that this equation has no solution for \(P(+|BB = 1) \leq 1\). This means that for a 5% false positive rate and for the given incidence \(P(BB = 1)\) it is impossible to obtain a test that is 99% reliable. Therefore in order to achieve 99% reliability, the false positive rate, \(P(+|BB = 0)\), has to be reduced, as well.

(d) Let us denote by \(x = P(+|BB = 1) = P(−|BB = 0)\) the reliability of the test. Then requiring a value of 99% for \(P(BB = 1|+)\) amounts to solving for \(x\) the following equation:

\[
0.99 = P(BB = 1|+) = \frac{xP(BB = 1)}{xP(BB = 1) + (1 - x)P(BB = 0)}
\] (11)

where \(P(BB = 1) = 10^{-6}\) and \(P(BB = 0) = 1 - 10^{-6} \approx 1.\) This gives for \(x\)

\[
x \approx \frac{1}{1 + 10^{-8}},
\] (12)

which means that the the reliability of the test ought to be in excess of 1 in \(10^8\). This is obviously not feasible and hence it is important to screen people before administering the test, i.e., to only test people who already show symptoms of the condition.

**CW1/3. Statistics in the kitchen.**

Before you toss the mushrooms raviolo in the pan, the probability of picking a mushrooms raviolo is 9/10 (i.e., probability as frequency). This becomes 10/11 after throwing in the extra mushrooms one. After you picked a mushrooms raviolo, the pan again contains 9 mushrooms ravioli and 1 ricotta raviolo (and you know this for a fact), hence the probability of picking mushrooms for the next pick is again 9/10 (events are independent).

In the second case, you have only one raviolo in your plate to start with, but its content is uncertain to you (i.e., probability here is best understood as degree of information, that’s to say, in the Bayesian framework). Since you have no other information, the principle of indifference states that you should give equal probability to the possible alternatives, so you assign a prior probability to the raviolo in your plate (denote by the symbol \(yr\), and \(m\) denotes mushrooms while \(r\) denotes ricotta): \(P(yr = m) = P(yr = r) = 1/2.\) After you have picked the raviolo and found it to be mushrooms \((m = 1)\) your posterior probability for the remaining raviolo in your plate being mushrooms is:

\[
P(yr = m|m = 1) = \frac{P(m = 1|yr = m)P(yr = m)}{P(m = 1|yr = m)P(yr = m) + P(m = 1|yr = r)P(yr = r)}
\] (13)

\[
= \frac{1}{1 + \frac{P(m = 1|yr = r)}{P(m = 1|yr = m)P(yr = m)}} = \frac{1}{1 + \frac{1}{2}} = \frac{2}{3}.
\]

Notice that here we have used that \(P(m = 1|yr = r) = 1/2\) (if the raviolo originally in your plate is ricotta, you have 50% probability of picking mushrooms) and \(P(m = 1|yr = m) = 1\) (if the raviolo originally in your plate is mushrooms, you are certain that you would pick mushrooms).
The same conclusion can be reached by enumerating all the possible outcomes of the experiment (e.g., using a tree diagram): (first raviolo = m, 2nd raviolo = m), (first raviolo = m, 2nd raviolo = m, inverting the two ravioli), (first raviolo = r, 2nd raviolo = m), (first raviolo = m, 2nd raviolo = r). Notice that (first raviolo = r, 2nd raviolo = r) is impossible, for you know that your friend’s raviolo is mushrooms. Of the 4 possible outcomes, the possibility (first raviolo = r, 2nd raviolo = m) is excluded, because your first raviolo is mushrooms, not ricotta. Therefore there are only 3 possible outcomes compatible with the observation. Of those 3, 2 have that the second raviolo = m. Therefore the probability of the remaining raviolo being mushrooms is 2/3.
CW2/1. Coin tossing. The solutions are:

(a) The likelihood function is given by

\[ \mathcal{L}(p) = P(r = H | p, n) = \binom{n}{H} p^H (1 - p)^{n-H}, \]

where the unknown parameter is \( p \) and the data are the number of heads, \( H \) (for a fixed number of trials, \( n = 10 \) here).

(b) The Maximum Likelihood Estimator (MLE) for the success probability \( p \) is found by maximising the log likelihood, see Example 20 in the handout:

\[ \frac{\partial \ln \mathcal{L}(p)}{\partial p} = \frac{\partial}{\partial p} \left( \ln \binom{n}{H} + H \ln p + (n - H) \ln(1 - p) \right) = \frac{H}{p} - \frac{n - H}{1 - p} = 0 \]

\[ \iff \quad p_{ML} = \frac{H}{n}. \]

Therefore the ML value for \( p \) is \( p_{ML} = 0.8 \).

(c) We approximate the likelihood function as a Gaussian, with standard deviation given by minus the curvature of the log-likelihood at the peak:

\[ \mathcal{L}(p) \approx \mathcal{L}_{\text{max}} \exp \left( -\frac{1}{2} \frac{(p_{ML} - p)^2}{\Sigma^2} \right), \]

where (see Eq. (65) in the handout)

\[ \Sigma^{-2} = -\left. \frac{\partial^2 \ln \mathcal{L}(p)}{\partial p^2} \right|_{p=p_{ML}} = \left. -\frac{\partial}{\partial p} \left( \frac{H}{p} - \frac{n - H}{1 - p} \right) \right|_{p=p_{ML}} = \frac{H - 2Hp + p^2n}{p^2(1 - p)^2} \left|_{p=p_{ML}} = \frac{n}{H} \left( 1 - \frac{H}{n} \right). \]

The 1σ confidence interval for \( p \) is given by \( \Sigma = 0.13 \). Therefore the result would be reported as \( p = 0.80 \pm 0.13 \).

(d) Following the hint, the number of σ confidence with which the hypothesis that the coin is fair can be ruled out is given by

\[ \frac{|p_{ML} - \frac{1}{2}|}{\Sigma} = \frac{0.8 - 0.5}{0.13} = 2.31. \]

Therefore the fairness hypothesis can be ruled out at the \( \sim 2.3 \sigma \) level.

(e) Using above equations, the MLE for the success probability is still \( p_{ML} = 0.8 \), as before. However, the uncertainty is now much reduced, because of the large number of trials. In fact, we get \( \Sigma = 0.013 \) (notice how the uncertainty has decreased by a factor of \( \sqrt{n} \), as expected. I.e., 100 times more trials correspond to a reduction in the uncertainty by a factor of 10). The fairness hypothesis can now be excluded with much higher confidence: of \( p = 1/2 \), expressed in number of sigmas:

\[ \text{number of sigmas} = \frac{|p_{ML} - \frac{1}{2}|}{\Sigma} = \frac{0.8 - 0.5}{0.013} = 23.1 \approx 23. \]

This constitutes more than decisive evidence against the hypothesis that the coin is fair. Notice however that the Gaussian approximation to the likelihood we employed will most probably not be accurate so far into the tails of the likelihood function (i.e., the Taylor expansion on which it is based is a local expansion around the peak).
CW2/2. **Counting experiment.**

(a) The discrete PMF for the number of counts $r$ of a Poisson process with average rate $\lambda$ is (assuming a unit time, $t = 1$ throughout)

$$P(r) = \frac{\lambda^r}{r!} e^{-\lambda}$$

(b) In this case

$$P(\hat{r}_i | \lambda) = \frac{\lambda^{\hat{r}_i}}{\hat{r}_i!} e^{-\lambda},$$

for each independent measurement $\hat{r}_i$. So the joint likelihood is given by (as measurements are independent)

$$\mathcal{L}(\lambda) = \prod_{i=1}^{M} P(\hat{r}_i | \lambda) = \prod_{i=1}^{M} \frac{\lambda^{\hat{r}_i}}{\hat{r}_i!} e^{-\lambda}.$$  \(7\)

(c) The Maximum Likelihood Principle states that the estimator for $\lambda$ can be derived by finding the maximum of the likelihood function. The maximum is found more easily by considering the log of the likelihood

$$\ln \mathcal{L}(\lambda) = \sum_{i=1}^{M} [\hat{r}_i \ln(\lambda) - \ln(\hat{r}_i!) - \lambda].$$

with the maximum given by the condition $d\ln \mathcal{L}/d\lambda = 0$.

We have

$$\frac{d\ln \mathcal{L}}{d\lambda} = \sum_{i=1}^{M} \left[ \frac{\hat{r}_i}{\lambda} - 1 \right]$$

$$= \frac{1}{\lambda} \sum_{i=1}^{M} \hat{r}_i - M.$$

So the Maximum Likelihood (ML) estimator for $\lambda$ is

$$\lambda_{ML} = \frac{1}{M} \sum_{i=1}^{M} \hat{r}_i,$$

which is just the average of the observed counts.

(d) The Taylor expansion is (see Eq. (63) in the handout)

$$\ln \mathcal{L}(\lambda) = \ln \mathcal{L}(\lambda_{ML}) + \frac{d\ln \mathcal{L}}{d\lambda} \bigg|_{\lambda=\lambda_{ML}} (\lambda - \lambda_{ML}) + \frac{1}{2} \frac{d^2\ln \mathcal{L}}{d\lambda^2} \bigg|_{\lambda=\lambda_{ML}} (\lambda - \lambda_{ML})^2 + \ldots.$$  

By definition the linear term vanishes at the maximum so we just need the curvature around the ML point

$$\frac{d^2\ln \mathcal{L}}{d\lambda^2} = -\sum_{i=1}^{M} \frac{\hat{r}_i}{\lambda^2},$$

such that

$$\frac{d^2\ln \mathcal{L}}{d\lambda^2} \bigg|_{\lambda=\lambda_{ML}} = \frac{1}{\lambda_{ML}^2} \sum_{i=1}^{M} \hat{r}_i = -\frac{M}{\lambda_{ML}} = -\frac{M}{\lambda_{ML}}.$$
Putting this into the Taylor expansion gives
\[ \ln L(\lambda) = \ln L(\lambda_{ML}) - \frac{1}{2} \frac{M}{\lambda_{ML}} (\lambda - \lambda_{ML})^2, \]
which gives an approximation of the likelihood function around the ML point
\[ L(\lambda) \approx L_0 \exp\left( -\frac{1}{2} \frac{M}{\lambda_{ML}} (\lambda - \lambda_{ML})^2 \right), \]
(the normalisation constant \( L_0 \) is irrelevant).

So the likelihood is approximated by a Gaussian with variance
\[ \Sigma^2 = \frac{\lambda_{ML}}{M}. \]

Comparing this with the standard result for the variance of the mean for the Gaussian case, i.e.
\[ \Sigma^2 = \frac{\sigma^2}{M}, \]
where \( M \) is the number of measurements and \( \sigma \) is the standard deviation of each measurement, we can conclude that the variance of the Poisson distribution itself is indeed
\[ \sigma^2 = \lambda. \]

**CW2/3. Gaussian measurements with variable variance.**

(a) The photon counts follow a Poisson distribution. We know that the MLE for the Poisson distribution is the observed number of counts \( \hat{n}_i \) and its standard deviation is \( \sqrt{\hat{n}} \). However, for large \( n \) \((\gg 20)\) the Poisson distribution is well approximated by a Gaussian of mean \( \hat{n} \) and standard deviation \( \sqrt{\hat{n}} \). In this case, \( n \) is of order \( 10^5 \), hence the standard deviation intrinsic to the Poisson process (the so-called “shot noise”) is of order \( \sqrt{10^5} \approx 3 \cdot 10^2 \). The quoted experimental uncertainty is much larger than that (of order \( 10^4 \) for each datum), hence we can conclude that the statistical error is dominated by the noise in the detector rather than by the Poisson variance.

Therefore we can approximate the likelihood for each observation as a Gaussian with mean given by the observed counts \( \hat{n}_i \) and standard deviation given by the quoted error, \( \hat{\sigma}_i \):
\[
L_i(F) = \frac{1}{\sqrt{2\pi\hat{\sigma}_i}} \exp\left( -\frac{1}{2} \frac{(F - \hat{n}_i)^2}{\hat{\sigma}_i^2} \right) \quad (i = 1, \ldots, 4). \tag{8}
\]

(b) Since the measurements are independent, the total likelihood is the product of the 4 terms:
\[
L(F) = \prod_{i=1}^{4} L_i(F). \tag{9}
\]

(c) To estimate the mean of the distribution, we apply the MLE procedure for the mean \( \hat{F} \), obtaining:
\[
\frac{\partial \ln L(F)}{\partial \hat{F}} = -\sum_i \frac{F - \hat{n}_i}{\hat{\sigma}_i^2} = 0 \tag{10}
\]
\[\Leftrightarrow \hat{F}_{ML} = \sum_i \frac{\hat{n}_i}{\hat{\sigma}_i^2/\sigma^2}, \]
where
\[
\frac{1}{\sigma^2} \equiv \sum_i \frac{1}{\hat{\sigma}_i^2}. \tag{11}
\]
We thus see that the ML estimate for the mean is the mean of the observed counts weighted by the inverse error on each on them (verify that Eq. (10) reverts to the usual expression for the sample mean for $\hat{\sigma}_i = \hat{\sigma}$ for $(i = 1, \ldots, 4)$, i.e., if all observations have the same error). This automatically gives more weight to observations with a smaller error.

(d) From the given observations, one thus obtains $F_{\text{ML}} = 29.2 \times 10^4$ photons/cm$^2$. By comparison the sample mean is $\bar{F} = 30.3 \times 10^4$ photons/cm$^2$.

(e) The inverse variance of the mean is given by the second derivative of the log-likelihood evaluated at the ML estimate, see Eq. (65) in the handout:

$$\Sigma^{-2} = -\frac{\partial^2 \ln \mathcal{L}(F)}{\partial F^2} \bigg|_{F = F_{\text{ML}}} = \sum \frac{1}{\hat{\sigma}_i^2}. \quad (12)$$

(again, it is simple to verify that the above formula reverts to the usual $N/\hat{\sigma}^2$ expression if all measurements have the same error).

Therefore the variance of the mean is given by $\Sigma^2 = 0.46 \times 10^8$ (photons/cm$^2$)$^2$, and the standard deviation is $\Sigma = 0.7 \times 10^4$ photons/cm$^2$. Our measurement can thus be summarized as $F = (29.2 \pm 0.7) \times 10^4$ photons/cm$^2$. 
