Poisson approximation by coupling

For those interested, this is the proof of a precise inequality for the approximation of a binomial using a Poisson distribution. (No need to study this for exams)

Approximation of a binomial using a Poisson distribution. Suppose that $S \sim \mathrm{Bin}(n,p)$ and $Y \sim \mathcal{P}(np)$ (i.e., $\lambda = np$). Then, for any subset A of $\{0,1,2,\ldots\}$,

$$|P(S \in A) - P(Y \in A)| \le np^2 = \lambda^2/n.$$

This is a special case of the following more general statement.

Approximation by a Poisson distribution Let $(X_i)_1^n$ be independent Bernoulli random variables with parameters p_i. Let Y_i be independent Poisson random variables with parameters p_i. Let $X = \sum_1^n X_i$ and $Y = \sum_1^n Y_i$. Then

$$|P(X \in A) - P(Y \in A)| \le \sum_1^n p_i^2.$$

The first statement is a consequence of the second because a binomial (n,p) can be viewed as a sum of n independent Bernoulli p and a Poisson np can be view as the sum of n independent Poisson p. This last fact can be seen as follows. Let U, V be independent Poisson with respective parameters u, v defined on a same probability space (Ω, P), and let $Z = U + V$. Then

$$\begin{aligned} P(Z = k) &= \sum_{\ell=0}^{k} P(U = \ell, V = k - \ell) \\ &= \sum_0^k e^{-u} \frac{u^\ell}{\ell!} e^{-v} \frac{v^{k-\ell}}{(k-\ell)!} \\ &= e^{-(u+v)} \frac{(u+v)^k}{k!}. \end{aligned}$$

This shows that the sum of two independent Poisson random variables is a Poisson random variable (with parameter given by the sum of the two original parameters). Induction allows us to extend the statement to more than two independent Poisson random variables. It is now clear the first statement follows from the second by taking all the p_is equal to p.

Proof of the second statement: Coupling Technique

This is an interesting and instructive proof if you are interested in probability theory beyond the scope of this course.
In order to prove the second statement, we are going to take advantage of the fact that it is enough to prove that statement for some particular collections of X_is and Y_is
because the statement is about the {\em distributions} of X and Y, not the random variable themselves.

We assume we are given the random variables $Y_i, 1 \le i \le n$, where Y_i is Poisson with parameter p_i. We then consider a sequence of Bernoulli random variables $U_i, 1 \le i \le n$, with parameters

$$P(U_i = 1) = 1 - (1 - p_i)e^{p_i},$$

and which are mutually independent and mutually independent of the Y_i (all these random variables live on the same probability space (Ω, P)).
The formula for the parameter of U_i is chosen so that

$$P(U_i = 0, Y_i = 0) = (1 - p_i)e^{p_i} e^{-p_i} = 1 - p_i$$

is the same as the probability of 0 for a Bernoulli with parameter p_i.

Observe that $(1 - p_i)e^{p_i} \le 1$ because $1 - x \le e^{-x}$ ($y = 1 - x$ is the equation of the tangent to $y = e^{-x}$ at the point $(0, 1)$ and this tangent is below the curve, e.g., because $x \mapsto e^{-x}$ is a convex function). This inequality is also the same as $1 - e^{-x} \le x$ which we will use below.

Next, we use the Y_is and U_is to construct a special sequence of independent Bernoulli random variables X_is with parameter p_is. Each X_i is constructed using Y_i and \$$U_i$ only and that is why the X_is are mutually independent. The X_is will NOT be independent of the Y_is. Indeed, we set

$$X_i = \begin{cases} 0 & \text{if } Y_i = U_i = 0 \\ 1 & \text{otherwise.} \end{cases}$$

Obviously $P(X_i = 0) = 1 - p_i$ as mentioned above and it follows that X_i is Bernoulli p_i.

Next we compute the probability that $X_i \neq Y_i$.

$$\begin{aligned} P(X_i \neq Y_i) &= P(X_i = 1, Y_i \neq 1) = P(X_i = 1, Y_i = 0) + P(Y_i > 1) \\ &= P(U_i = 1, Y_i = 0) + P(Y_i > 1). \end{aligned}$$

({\em Why is it true that $P(X_i = 1, Y_i = 0) = P(U_i = 1, Y_i = 0)$?})

Now we compute $P(Y_i > 1) = 1 - e^{-p_i}(1+p_i)$ and

$$P(U_i = 1, Y_i = 0) = e^{-p_i}(1 - (1-p_i)e^{p_i}) = e^{-p_i} - (1-p_i).$$

It follows that

$$P(U_i = 1, Y_i = 0) + P(Y_i > 1) = p_i - p_i e^{-p_i} = p_i(1 - e^{-p_i}) \leq p_i^2.$$

The next step is to form

$$X = \sum_1^n X_i, \quad Y = \sum_1^n Y_i.$$

If $X \neq Y$ it must be that $X_i \neq Y_i$ for at least one i.
It follows that

$$P(X \neq Y) \leq \sum_1^n P(X_i \neq Y_i) \leq \sum_1^n p_i^2.$$

Observe that

$$P(X \in A) = \sum_{\ell \in A} \sum_k P(X = \ell, Y = k) \text{ and } \quad P(Y \in A) = \sum_{\ell} \sum_{k \in A} P(X = \ell, Y = k).$$

So

$$\begin{aligned} |P(X \in A) - P(Y \in A)| &= \left| \sum_{\ell \in A} \sum_k P(X = \ell, Y = k) - \sum_{\ell} \sum_{k \in A} P(X = \ell, Y = k) \right| \\ &= \sum_{(\ell,k): \ell \in A, k \notin A \text{ or } \ell \notin A, k \in A} P(X = \ell, Y = k) \\ &\leq \sum_{(\ell,k): \ell \neq k} P(X = \ell, Y = k) = P(X \neq Y) \leq \sum_1^n p_i^2. \end{aligned}$$

The proof is done.